

Human Nature and the World around Us: Three Essays in Economics

Dissertation
submitted to the
Faculty of Business, Economics and Informatics
of the University of Zurich

to obtain the degree of
Doktor der Wirtschaftswissenschaften, Dr. oec.
(corresponds to Doctor of Philosophy, PhD)

presented by

Christian Zünd
from Zurich, ZH

approved in April 2020 at the request of

Prof. Dr. Ernst Fehr
Prof. Dr. Michel Maréchal
Prof. Dr. Pietro Biroli

The Faculty of Business, Economics and Informatics of the University of Zurich hereby authorizes the printing of this dissertation, without indicating an opinion of the views expressed in the work.

Zurich, 01.04.2020

The Chairman of the Doctoral Board: Prof. Dr. Steven Ongena

For Corina,

on the fourth

anniversary of

our marriage.

Acknowledgements

I would like to thank my supervisor Ernst Fehr for his guidance and support, for his commitment to Economics as a comprehensive social science, and for creating a small research paradise in Zurich. There are very few other places that grant their PhD students such freedom to pursue exciting projects, and I am grateful for the opportunity to conduct my doctoral research at this department.

Special thanks belongs to the other members of my dissertation committee. When Michel Maréchal asked me to join Project Honesty in my first year, neither of us would have guessed just how much time and energy we would have to invest, nor that five years later our work would be featured by such esteemed media outlets as ARD's *Wer weiss denn sowas?* I am also indebted to Pietro Biroli for his contagious enthusiasm and for inducing me into the social science genetics community.

Furthermore, I benefited tremendously from collaborating with my many brilliant coauthors: Alain Cohn and, later, David Tannenbaum from Project Honesty; Leandro Carvalho and Job Harms from the Postcode Lottery; Sonja Vogt, Charles Efferson, and Flávio Eiró from my Brazilian endeavors; and the dozens of people from the Risk GWAS, in particular, Richard Karlsson Linnér, Jonathan Beauchamp, Dan Benjamin, and David Cesarini. I would also like to thank the friends, office mates, and fellow-travelers from the PhD program.

I would like to extend a heartfelt thank you to the fantastic administrative staff at the Department of Economics, especially Sally Gschwend, Helen Bernhard, Mirjam Britschgi, and Karin Wyss. I lost count of how many times their uncomplicated helpfulness saved my day. I am also thankful for the support of our IT engineering team, especially Adrian Etter and Andi Saurer, whose skilled eyes spared me countless hours of debugging my code.

I am thankful for the financial support from the Excellence Foundation, the Swiss National Science Fund, and the Department of Economics. I thank Johannes Abeler and Ray Duch for hosting me during my research stay at the University of Oxford and all the biologists who let me into their labs and helped to bring me up to speed on genetics.

Finally, I am deeply grateful for my friends and family. I thank my parents Urs and Regina for the countless ways in which they aided me—from the genes they endowed in me at the very beginning, their generous personal and financial support, to their company for many years to come. My brother Michael deserves mentioning for helping me to count and package thousands

of keys and plastic wallets while watching the Simpsons.

Most of all, I would like to thank Corina, my wonderful wife. Listing all the practical ways by which she contributed to the completion of my dissertation would not do justice to my profound gratefulness. Thank you for believing in me and for the audacity to marry a perpetual student. Four years ago, to the day.

Christian Zünd, January 16, 2020

Contents

Acknowledgements	v
1 Dissertation Overview	1
2 Civic Honesty around the Globe	13
I. Introduction	14
II. Field Experiment	14
A. Design	15
B. Results	15
III. Mechanism	17
IV. Prediction Experiments	19
V. Conclusion	20
3 Local Corruption, Income Underreporting, and Policy Effectiveness	23
I. Introduction	24
II. Institutional Background	29
A. Bolsa Família’s Incentives for Educational Participation	29
B. Bolsa Família’s Beneficiary Selection Process	31
C. Bolsa Família’s Safeguards against Corruption	32
D. Bolsa Família’s Vulnerabilities to Corruption	34
III. Estimating the Effectiveness of Bolsa Família	35
A. Data—the Cadastro Único	35
B. Sample—Randomly Admitted Families	36
C. Estimating the Treatment Effects	37
D. Validating the Identification Strategy	38
E. Bolsa Família Increases School Enrollment	39
IV. Local Corruption Affects Program Effectiveness	41
A. Random Audits as Exogenous Shocks to Corruption	41
B. Validating the Corruption Reduction after Random Audits	43
C. Estimating the Change in Treatment Effects	46
D. Bolsa Família Is More Effective after a Random Audit	46

V.	Understanding the Mechanism	50
A.	Income Underreporting and Mistargeting	50
B.	A Model of Income Underreporting	51
C.	Testing the Model in the Administrative Data	55
D.	Testing the Model in a Field Experiment	58
E.	Social Norms on Income Underreporting	62
VI.	Financial Gains from Bolsa Família	64
A.	Retaining Benefit Cards	64
B.	Self-Registration in the Cadastro Único	65
C.	Embezzling Complementary Funds	66
VII.	Alternative Explanations	66
A.	School Monitoring	66
B.	Data Inconsistencies	67
C.	CRAS Processes	68
D.	CRAS Infrastructure and Funding	69
E.	Complementary Actions and Programs	70
F.	Governance and Oversight	71
G.	Social Control	72
VIII.	Conclusion	73
4	Gene-Environment Interplay and Alcohol Licensing Policy	81
I.	Introduction	82
II.	Empirical Methods	84
A.	The UK Biobank	84
B.	Alcohol Availability	85
C.	Local Licensing Policy	87
D.	Polygenic Scores (PGS)	88
E.	Empirical Specification	90
III.	Results	92
A.	Genetic Predisposition and Alcohol Consumption	92
B.	Genetic Predisposition and Self-Selection	94
C.	Genetic Predisposition and Elasticity of Demand	96
D.	Genetic Predisposition and Licensing Policy	98
E.	Genetic Predisposition and Public Health	101
IV.	Conclusion	103
A2	Appendices to: Civic Honesty around the Globe	109
I.	Materials and Methods	110
A.	Lost Wallet Experiments	110

B.	Selection of Countries and Cities	110
C.	Number of Drop-Offs	110
D.	Selection of Drop-Off Locations	111
E.	The Wallets	112
F.	Drop-Off Procedure	113
G.	Experimental Conditions	114
H.	Measuring Civic Honesty	114
I.	Measuring Recipient Characteristics and Situational Factors	116
J.	Country-Level Correlates of Civic Honesty	117
K.	Survey Experiments	139
L.	Prediction Study: Non-Expert Sample	142
M.	Prediction Study: Expert Sample	143
II.	A Conceptual Framework for Civic Honesty	144
III.	Results	148
A.	Civic Honesty across Countries	148
B.	Civic Honesty under High Stakes	150
C.	Testing for Altruism	151
D.	Evidence for Theft Aversion	152
E.	Prediction Data: Non-Expert Sample	156
F.	Prediction Data: Expert Sample	159
G.	Cross-Country Correlates of Civic Honesty	160
IV.	Alternative Explanations	173
A.	Fear of Punishment	173
B.	Returning the Wallet but Pocketing the Money	176
C.	Possible Finder's Fee for Returning a Wallet	177
V.	Robustness Checks	179
A.	Individual and Situational Factors	179
B.	Experimenter Effects	179
C.	Differences in Email Usage	185
D.	Differences in Economic Development	185
A3 Appendices to: Local Corruption, Income Underreporting, and Policy Eff.		195
I.	Derivation of the Registration Model	196
A.	Properties of the Cumulative Binomial Distribution	196
B.	Deriving the Reporting Function	196
C.	Optimality of the Reporting Function	199
D.	Implications of the Reporting Function	200
II.	Robustness	202

A.	Following Children for More Than Two Years	202
B.	Non-Parametric Controls for Age and Gender	203
C.	Including Teenagers	204
D.	Only Original Random Audit Program	205
E.	Including Audits in the Same Year	206
F.	Updating the Cadastro Único	207
G.	Including Only Complete Years	208
H.	Treatment Propensity	209
III.	Details and Results of the Field Experiment	212
A.	Sample	212
B.	Treatments	212
C.	Sending the Emails	212
D.	Response Times	214
E.	Analysis of Email Content	214
F.	Low Response Rate	216
G.	Robustness	217
IV.	Details and Results of the Social Norms Experiment	222
A.	Recruitment and Sample	222
B.	Experimental Design	224
C.	Social Norms	224
D.	Beliefs	226
E.	Rule-Following	227
V.	Additional Figures	229
VI.	Additional Tables	238
A4	<i>Appendices to: Gene-Environment Interplay and Alcohol Licensing Policy</i>	251
I.	Additional Figures	252
II.	Additional Tables	260
	Curriculum Vitae	271

List of Figures

2.1	Share of Wallets Reported in the NoMoney and Money Condition by Country	16
2.2	Reporting Rates as a Function of Monetary Stakes	17
2.3	Actual versus Predicted Reporting Rates	20
3.1	Timeline of Random Audits	43
3.2	Bolsa Família Is More Effective after a Random Audit	47
3.3	Reported Income as a Function of True Income	53
3.4	Predicted Treatment Effects for Low and High Rates of Underreporting	54
3.5	Program Effectiveness by Income Bracket	57
3.6	Response Rate to Requests from Eligible and Ineligible Families	59
3.7	Social Norms Don't Change after a Random Audit	63
4.1	Distribution of the Alcohol Availability Measures	86
4.2	Distribution and Predictive Power of the Polygenic Score	90
4.3	Relationship between the Polygenic Score and Drinking Behavior	93
4.4	Evidence of Gene-Environment Correlation	95
4.5	Drinking Behavior and Alcohol Availability by Polygenic Score	97
A2.1	Example Lost Wallet	113
A2.2	Cumulative Distribution Function of Response Times	115
A2.3	Example Survey Scenario	140
A2.4	Response Patterns as a Function of Altruism (α) and Theft Aversion (γ) . . .	146
A2.5	Correlates of Civic Honesty	162
A2.6	Explaining Cross-Country Variation	166
A2.7	Correlates of Civic Honesty: Original Data Only	168
A2.8	Correlates of Civic Honesty: NoMoney	169
A2.9	Correlates of Civic Honesty: Money	170
A2.10	Correlates of Civic Honesty: Controlling for Geography	171
A2.11	Correlates of Civic Honesty: Controlling for Country GDP	172
A2.12	Regression-Adjusted Ranking	180
A2.13	Country Ranking for Hotels	186
A2.14	Regression-Adjusted Ranking: Email Usage	187

A2.15	Regression-Adjusted Ranking: Country GDP	188
A3.1	Overlap of Treatment Propensities	209
A3.2	Email Delivery Rates by Wave and Previous Audit	213
A3.3	Response Times by Treatment and Previous Audit	214
A3.4	Facebook Recruitment Advert (English Translation)	222
A3.5	Distribution of Reported Winning Coin Flips in the Honesty Game	227
A3.6	Cadastral Único Registration Rate	229
A3.7	Geographic Coverage of Marginal Priority Strata	229
A3.8	Distribution of Bolsa Família's Effectiveness in Different Municipalities	230
A3.9	Distribution of Random Audits	231
A3.10	Distribution of the "Past Audit" Indicator	232
A3.11	Financial Loss Uncovered by the Random Audits	233
A3.12	Rates of Illegitimate Payments for Citizens and Public Servants	233
A3.13	School Enrollment and Family Income	234
A3.14	Change in Exclusions after a Random Audit	234
A3.15	CRAS Infrastructure Doesn't Change after a Random Audit	235
A3.16	Complementary Actions Don't Change after a Random Audit	236
A3.17	Governance of Social Programs Doesn't Change after a Random Audit	237
A4.1	Geographic Distribution of UK Biobank Participants	252
A4.2	Geographic Distribution of Pubs in the United Kingdom	253
A4.3	Predictive Power of the Polygenic Score	254
A4.4	Local Licensing Policy Is Uncorrelated with Genetic Predisposition	255
A4.5	Participants with a High PGS Have Less Elastic Demand (Retailers)	256
A4.6	High PGS Participants React Less to Policy (Licensed Premises)	257
A4.7	High PGS Participants React Less to Policy (No-Sale Licenses)	258
A4.8	High PGS Participants React Less to Policy (24h Supermarkets)	259

List of Tables

3.1	Balancedness of Family Characteristics in Marginal Priority Strata	39
3.2	Bolsa Família Increases School Enrollment	40
3.3	Random Audits Significantly Decrease Corruption	45
3.4	Robustness to Alternative Specifications	48
3.5	Underreporting Decreases after Random Audits	56
3.6	Email Text by Treatment	59
3.7	Response Rates to Requests from Eligible and Ineligible Families	61
3.8	Vignettes and Elicited Social Norms	63
4.1	Self-Selection into Areas with Higher Alcohol Availability	96
4.2	Gene-Environment Interaction for Licensing Policy (Licensed Premises)	99
4.3	Gene-Environment Interaction for Licensing Policy (No-Sales Licenses)	100
4.4	Gene-Environment Interaction for Licensing Policy (24h Supermarkets)	101
A2.1	Sample Overview	121
A2.2	Descriptive Statistics and Randomization Check for the United Kingdom . . .	133
A2.3	Descriptive Statistics and Randomization Check for Poland	134
A2.4	Descriptive Statistics and Randomization Check for the United States	135
A2.5	Descriptive Statistics and Randomization Check for the Global Data	136
A2.6	Analysis of Rejections	137
A2.7	Analysis of Response Times	138
A2.8	Reporting Rates in the Money and NoMoney Condition	149
A2.9	Reporting Rates in NoMoney, Money, and BigMoney Condition	150
A2.10	Reporting Rates in Money-NoKey Condition	151
A2.11	Survey Responses across Experimental Conditions	153
A2.12	Survey Responses across Experimental Conditions, Full Sample	154
A2.13	Predictions of Reporting Rates across Experimental Conditions	158
A2.14	Civic Honesty and Lost Property Laws	174
A2.15	Civic Honesty and Presence of Security Cameras	175
A2.16	Civic Honesty and Social Monitoring	176
A2.17	Civic Honesty and Beliefs about Finder's Fees	178

A2.18	Drop-Offs by Experimenters and Country: Overlaps	182
A2.19	Experimenter Effects	183
A2.20	Experimenter Effects (continued)	184
A3.1	Robustness: Following Children for up to Three Years	202
A3.2	Robustness: Non-Parametric Controls for Age and Gender	203
A3.3	Robustness: Including Teenagers	204
A3.4	Robustness: Only Original Random Audit Program	205
A3.5	Robustness: Including Audits in the Same Year	206
A3.6	Robustness: Excluding Families with Old Data	207
A3.7	Robustness: Excluding 2017	208
A3.8	Robustness: Inverse Probability Weights	210
A3.9	Robustness: Families with Treatment Propensity 10–90%	211
A3.10	Analysis of Email Content (Heckman Selection Model)	215
A3.11	Robustness: Including Undelivered Emails	218
A3.12	Robustness: First Wave Only	219
A3.13	Robustness: Separate Control Treatments	220
A3.14	Balancedness of Individual Characteristics in Online Sample	223
A3.15	Social Norms Don’t Change after a Random Audit	225
A3.16	Beliefs Don’t Change after a Random Audit	226
A3.17	Bolsa Família Increases School Enrollment (No Child Fixed Effects)	238
A3.18	Corruption in Bolsa Família and the Educational System	239
A3.19	Treatment Effects for Families Registered at Home	240
A3.20	Change in Sanctions after a Random Audit	241
A3.21	The Bolsa Família Management Index Doesn’t Change...	242
A3.22	... but the Components Do.	243
A3.23	CRAS Centers Don’t Update the Cadastro Único Differently	244
A3.24	CRAS Centers Don’t Employ Different People	245
A3.25	Beneficiary Involvement at CRAS Doesn’t Change Much	246
A3.26	Municipalities Don’t Spend More on Social Assistance	247
A3.27	Municipalities Don’t Spend More on Education	248
A3.28	Whistleblowing Doesn’t Change after a Random Audit	249
A4.1	Summary Statistics of Drinking Outcomes in the UKB	260
A4.2	Correlation of Alcohol-Related Outcomes	261
A4.3	Local Licensing Policy Is Uncorrelated with Genetic Predisposition	262
A4.4	Alcohol Consumption as a Function of the PGS, Pubs, and Retailers	263
A4.5	Moving as a Function of the PGS and the Number of Pubs	264
A4.6	Moving as a Function of the PGS and the Number of Retailers	265

A4.7	Moving to Another Local Authority as a Function of the PGS and Licensing .	266
A4.8	Liver Diseases and Genetic Predisposition	267
A4.9	Alcohol-Related Mental and Behavioral Disorders and Genetic Predisposition	268
A4.10	Other Alcohol-Related Diseases and Genetic Predisposition	269

Chapter 1

Dissertation Overview

"... one does not argue about tastes for the same reason that one does not argue over the Rocky Mountains—both are there, will be there next year, too, and are the same to all men."

— *Stigler and Becker (1977, p.76)*

In "De Gustibus Non Est Disputandum" George Stigler and Gary Becker famously invoke the Rocky Mountains to illustrate not only the futility of quarreling over individual tastes but to argue more generally that assumptions of different or unstable tastes "really have only been ad hoc arguments that disguise analytical failure" (p.89). Be this as it may, it always struck me that their continued existence and their sameness unto all men are the least exciting qualities of the Rocky Mountains. Surely a geologist would agree that there are much more enlightening questions, such as: How did the Rocky Mountains come into existence? What environmental forces sculpted them to look the way they do? What is their impact on the climate of North America?

This dissertation encompasses three papers on the foundations of human nature and how they interact with the world around us and the institutions we create to shape our actions. Just as a geologist might survey the Rocky Mountains to understand whether their morphology in Canada similar to that of their southern ranges, Chapter 2 presents a survey of our tendency for honest behavior in 40 different countries. Where an earth scientist might ask whether plate tectonic movement during the Laramide orogeny has raised the cordillera uniformly over the entirety of the Rockies, we investigate whether civic honesty is equally affected by financial incentives in each of the countries in our sample. Using a large-scale field experiment to measure how likely a supposedly lost wallet is returned, the paper documents a strong universal tendency for honest behavior—in virtually all countries, wallets with cash were more likely to be reported. Nevertheless, there are considerable differences in reporting rates. Observed civic honesty correlates significantly with economic development and the quality of local institutions. It is also shaped in predictable ways by deep-rooted environmental factors such as a country's geography and climate, and by historical indicators of essential cultural dimensions. Thus, even though the near-ubiquity of civic honesty around the globe—the main finding of this chapter—is a powerful reminder of our shared human nature, the degree to which it is manifest in different places is moderated by the world around us.

While Chapter 2 offers a global perspective on civic behavior, Chapter 3 shows how individuals' honesty and local corruption jointly affect the effectiveness of government policy within a country. In much the same way as our geologist might inquire what underlies the strong Bouguer gravity anomalies in parts of the Colorado Rockies, this paper investigates why Brazil's Bolsa Família, a *centrally administrated* cash transfer program, performs significantly worse in municipalities with higher *local* corruption. In the paper, I exploit the program's central selection process to estimate the causal effect of Bolsa Família on children's school enrollment in different municipalities, and I then use a second natural experiment to show that local corruption

significantly reduces the program's impact. A closer examination of the mechanism reveals that the higher prevalence of families who misrepresent their income can account for the variation in program effectiveness. This difference in income underreporting allows for two interpretations. Local corruption could erode social norms and make residents less averse to dishonesty, or it could leave people's inclination to cheat unchanged and merely make it easier to get away with misreporting one's income. These two explanations are tested in an online and a field experiment, respectively, and the results strongly favor the latter. This again goes to show that the world around us influences to what extent our natural disposition to honesty is reflected in our behavior.

The findings from Chapters 2 and 3 are compatible with a Rocky Mountains interpretation of human nature: our preferences and tendencies are essentially the same for everyone; only the environment determines their expression. In contrast, Chapter 4 shows that some of our tastes are heterogeneous in profound ways. While these differences are not commonly observable, they can be uncovered through novel technologies emerging from the natural sciences. In the same manner as a geologist might use breakthroughs in mass spectrometry to unveil previously hidden mineralogic diversity in the Rocky Mountains, the chapter uses genetic data to expose fundamental and deep-seated inequality in individuals' predisposition to consume alcohol. The paper demonstrates how these genetic differences, fixed at conception, lead to heterogeneous behavior in a similar environment and to diverging reactions to the same alcohol licensing policy. While the other two chapters measure environmental influences at the level of municipalities or even countries, the final chapter uses a fine-grained measure of local alcohol availability within 1000m of an individual's place of residence, and a polygenic score constructed from their genetic data for the predisposition to drink. With these tools, it shows that people with a high genetic propensity to drink self-select into environments with easier access to alcohol, react less to changes in the density of sales points, and respond less to restrictive licensing. The findings demonstrate that the supply-focused licensing policy clashes with individual predispositions and exacerbates genetic inequality, illustrating how advances in the natural sciences can enhance our understanding of central questions in the social sciences. Thus, unlike the previous chapters, the final chapter shows that not only does the environment moderate the expression of intrinsic tendencies, but individual differences in genetic endowments also moderate our response to the world around us.

In the remainder of this dissertation overview, I will briefly summarize each of the chapters, before closing with a note on their shared methodological contributions.

Chapter 2, *Civic Honesty around the Globe*, reports the results of a large-scale field experiment on civic honesty that I conducted together with Alain Cohn and Michel Maréchal.

Honest behavior is a central feature of economic and social life (Arrow, 1972; Algan and Cahuc, 2013). Without honesty, promises are broken, contracts go unenforced, taxes remain

unpaid, and governments become corrupt. Such breaches of honesty are costly to individuals, organizations and entire societies. Although there is robust experimental literature on the conditions that give rise to honest behavior (e.g., Ellingsen and Johannesson, 2004; Mazar et al., 2008; Gächter and Schulz, 2016; Gneezy et al., 2018; Abeler et al., 2019), little is known about how material incentives impact civic honesty, particularly in field settings. Most of the experimental literature on honest behavior involves modest financial stakes, has been conducted in laboratory settings, and tends to rely on populations from Western, educated, industrialized, rich and democratic societies (Henrich et al., 2010).

Theories of honesty make different predictions about the role of material incentives. Classic economic models based on rational self-interest suggest that, all else equal, honest behavior will become less common as the material incentives for dishonesty increase (Becker, 1968). Models of human behavior that incorporate altruistic or other-regarding preferences also predict dishonesty to rise with increasing incentives, as self-interest virtually always dominates concerns for the welfare of others — we care about others but not as much as we care about ourselves (Fehr and Schmidt, 1999; Charness and Rabin, 2002; Fisman et al., 2007). As a result, self-interest will play an increasingly prominent role in behavior as the material incentives for dishonesty grow. Psychological models based on self-image maintenance predict that people will cheat for profit so long as their behavior does not require them to negatively update their self-concept (Mazar et al., 2008; Duval and Wicklund, 1972). However, it is unclear *ex ante* whether self-image concerns will become more or less important as the incentives for dishonesty increase, and what form that relationship will take.

In order to investigate this fundamental trade-off, we turned in more than 17,000 apparently lost wallets (containing business cards, a shopping list, some money, and a key) at the reception desks of public and private institutions in 40 countries. We varied the amount of money that the wallets contained, including either no money, about 13 USD, or almost 100 USD.

We find that, on average, people are *more* likely to return wallets when they face a stronger incentive to steal. Although there are substantial differences in civic honesty across countries, the likelihood of returning a wallet is significantly higher in the majority of countries when the wallet contains more money. This result is truly a global phenomenon—in virtually all countries wallets with money were more likely to be reported, and the same is true for different institutions (e.g., hotels, banks, police stations, etc.) and irrespective of the finder’s age and gender.

Despite this near universal tendency for honest behavior, it tends to take both non-experts and professional economists by surprise. We conducted two online experiments where we asked participants to predict the results of our study, one with a general online sample from Amazon Mechanical Turk and one with economists who rank in the top 5% of their profession according to RePEC. Both groups predicted incorrectly that rates of honesty decline as the financial stakes increase. Rather than being the result of basic ignorance or of a *déformation professionnelle*, this failure to appreciate the scope of civic honesty reflects a misplaced cynicism with regard to

the honesty of others.

To identify the motives to keep or return the wallet, we conducted additional treatments and nationally representative online experiments in three countries. Our results are consistent with theoretical models that incorporate altruism and self-image concerns, but also suggest modification in that non-pecuniary motivations directly interact with the material benefits gained from dishonest behavior. In the extensive appendix, we show that other motives such as finder's fees or the risk of detection have little explanatory power, and we build on a rich literature of cross-country studies to explain the observed reporting rates. Wallet reporting rates are positively correlated with GDP and negatively correlated with income inequality and various corruption measures. Further analysis suggests that economically favorable geographic conditions, inclusive political institutions, and cultural values that extend moral norms beyond one's in-group are also positively associated with civic honesty.

Understanding the relationship between civic honesty and material incentives is not only practically relevant, but also theoretically important. In particular, our results indicate that people not only consider what is right or wrong, but that their moral considerations are inherently linked to the material benefits of dishonesty. When people stand to heavily profit from engaging in dishonest behavior, the desire to cheat increases but so do the psychological costs of viewing oneself as a thief—and sometimes the latter will dominate the former.

Chapter 3, *Local Corruption, Income Underreporting, and Policy Effectiveness*, examines the interplay of local corruption and civic behavior. The paper demonstrates how corruption negatively impacts policy effectiveness in a program that is explicitly designed to be unsusceptible to bribery, clientelism, and the embezzlement of money: Brazil's Bolsa Família program. Bolsa Família pays a monthly benefit to approximately 14 million families provided that their children attend school sufficiently often. Importantly, the program transfers funds directly to beneficiaries so that local officials cannot embezzle the money. Moreover, local officials are also bypassed in deciding who benefits from Bolsa Família; a central anonymized process selects beneficiaries to prevent clientelism. However, even if effective safeguards against bribery, clientelism, and embezzlement are in place, government programs are not necessarily immune to the corrosive impact of corruption.

One possible explanation for the negative effect of corruption is that it enables families to underreport their income to gain or keep Bolsa Família; even if corrupt *officials* can not embezzle funds, local corruption can make it easier for *families* to cheat. However, although it is generally assumed that income underreporting decreases the effectiveness of programs like Bolsa Família, this assumption has not previously been tested. This paper provides causal evidence linking local corruption, families' ability to successfully underreport their income, and lower causal treatment effects of Bolsa Família on school enrollment.

The project uses data from the Cadastro Único, the official database of the program, which

enables a novel identification strategy to estimate the causal effect of Bolsa Família on school enrollment: I can reconstruct the algorithm’s priority strata to identify families on the margin of the program that were randomly included or not included in Bolsa Família. The identification strategy has several advantages: First, because it uses the same data as the selection algorithm, there are no unobservable differences that affect who gets included in the program. Secondly, it is robust to families’ strategic motives, and the estimated treatment effects are unbiased conditional on self-reported income. Finally, unlike approaches that use variation in the program’s rollout or geographic coverage, it can estimate the effectiveness of the program at the municipality level and in different years.

The first main result of the chapter is that Bolsa Família is less effective in municipalities with more corruption, even though local officials cannot divert money from the program and have no say in the selection of beneficiaries. Using Brazil’s initiative to randomly audit municipalities as an exogenous shock to corruption, I show that Bolsa Família’s effect on school enrollment increases by about a third after a random audit.

I then test whether income underreporting can account for this result. Because municipalities are responsible for registering potential beneficiaries, local corruption can play the most prominent role at the registration stage. I consider a simplified model of Bolsa Família’s registration process, where families decide what income to report, and the families with the lowest self-reported incomes are included in the program. As predicted by the registration model, fewer families claim to have zero income, and fewer families report an income that is eligible for Bolsa Família after a municipality has been audited at random. Moreover, the treatment effect increases most for families that are predicted to have the strongest incentives to misrepresent their incomes.

As direct evidence for the income underreporting explanation, I conducted a field experiment with 6,998 Bolsa Família registration centers. Fictitious applicants asked about the possibility of receiving Bolsa Família, and their characteristics were experimentally varied to make them eligible or ineligible while holding everything else constant. Consistent with the income underreporting mechanism, registration centers in audited municipalities are significantly less likely to engage with ineligible families and to incorrectly state that a sender’s income is compatible with Bolsa Família. Taken together, these findings suggest that the effectiveness gains of Bolsa Família are the result of improvements in the way the program is targeted to the families that benefit from it the most.

Ruling out other alternative explanations, I demonstrate that changes in relevant local social norms—about underreporting income, colluding with underreporters, and blowing the whistle—are unable explain the results. I use the Krupka and Weber (2013) norm elicitation task to show that social norms do not change significantly after the audits in an incentivized online experiment with low-income participants from 424 municipalities, some of which had been randomly selected for audits in the past. Nor can improved school attendance monitoring, ad-

ministrative processes, infrastructure, complementary programs, tighter governance, or increased whistleblowing account for the increase in Bolsa Família’s effectiveness after a municipality has been audited at random.

Taken together, these results suggest that income underreporting can explain how, despite the program’s safeguards, local corruption undermines the effectiveness of Bolsa Família.

Chapter 4, *Genes, Pubs, and Drinks: Gene-Environment Interplay and Alcohol Licensing Policy in the UK*, reports the results of joint work with Pietro Biroli that investigates how genetic predisposition and local alcohol licensing policy interact and jointly influence people’s alcohol consumption choices.

Debates about the relative influence of nature versus nurture on human behaviors are amongst the oldest in the social sciences (Mulcaster, 1582; Hume, 1748; Darwin, 1859; Freud, 1930). In recent decades, however, it has become increasingly clear that pitting nature against nurture should be relinquished in favor of a more systemic view that considers the complex interplay that may exist between people’s genetic makeup and the environment in which they develop (Hunter, 2005; Heckman, 2007).

Alcohol consumption is worth our attention because of its prevalence and its negative consequences: it is estimated to be the third leading cause of preventable death, it has been related to more than 60 medical conditions, and it accounts for a share of the global burden of disease comparable to those of tobacco or hypertension (Mokdad et al., 2004; Room et al., 2005). In other words, drinking alcohol is one of the leading health behaviors that contribute to increasing health inequality. But what is the origin of this inequality? While recent research shows that our genes affect how much alcohol we drink, it is unclear how our genetic propensity influences our reaction to changes in the availability of alcohol and what this implies for effective alcohol licensing policy.

Our paper shows that the genetic propensity to drink alcohol contributes to health inequalities in two ways: by promoting selection into unfavorable environments, and by decreasing susceptibility to more restrictive licensing policies. Both negative selection and decreased susceptibility lead to higher alcohol intake and, eventually, alcohol-related diseases.

We combine genetic information from the UK Biobank with geo-coded locations of pubs and retailers, as well as data on alcohol licensing from local authorities in England and Wales. Using information on 700,000 genetic variants, we estimate a polygenic score that proxies the individual genetic propensity for alcohol consumption, and we use the coordinates of all pubs and the major retailers in the UK to construct a fine-grained measure of local alcohol availability.

Our results show that both living in proximity to many alcohol sales points and a high polygenic propensity for alcohol consumption are related to several drinking-related behaviors. Individuals with a high genetic propensity to drink also react less to changes in the availability of alcohol. Moreover, we find that individuals with a high polygenic risk self-select into envi-

ronments with greater alcohol availability, leading to substantial gene-environment correlation where carriers of similar genetic variants tend to cluster in the same area.

Turning to licensing policy, we show that individuals with a high polygenic risk respond less to restrictive licensing. Importantly, local licensing committees in England and Wales are not allowed to consider public health, much less the genetic predispositions of their residents, which mitigates concerns of reverse causality. We estimate the effect of licensing policy on alcohol consumption for individuals with low and high polygenic risk and find that a more restrictive licensing policy leads to decreased alcohol intake on average, but individuals with high polygenic risk are less responsive to policy change. Therefore, this public policy limiting the supply of alcohol tends to amplify existing genetic inequalities: it is more effective for those individuals who already have a low genetic predisposition to drinking, but it has less bite for those individuals who might need it the most.

Finally, using information on physician-diagnosed medical conditions from the National Health Service, we investigate the implication of our results from a public health perspective. Our results show that individuals with a high polygenic risk are significantly more likely to have an alcohol-related condition, including liver disease, psychological disorders, and various afflictions of the digestive system. These results hold even if we control for self-reported alcohol consumption, illustrating how incorporating genetic information can help us uncover relevant dimensions of health-inequality that are not commonly observable.

The results of the paper demonstrate how genetic information can shed light on the determinants and the dynamics of health inequalities, and how genetic endowments interact with individual choices and public health policy. We show that the effectiveness of supply-focused licensing policy as a tool to mitigate alcohol abuse can clash with individual predispositions and might actually exacerbate genetic inequality, suggesting the need for a more targeted approach.

Together, the three chapters showcase the potential of an integrated, methodologically diverse empirical approach. Chapter 2 complements a field experiment with nationally representative online surveys and extensive cross-country data from institutional, cultural, and political economics. While the field experiment provides external validity and global portability—a reality check for theories about honesty—, incorporating the online surveys enables us to deepen our understanding of the psychological underpinnings. Similarly, the extensive administrative database and the natural experiments used in Chapter 3 are ideally suited to prove the detrimental effect of corruption on the Bolsa Família program. Once the main result is established, the field and online experiments offer tailor-made tests to isolate income underreporting as the mechanism. Finally, Chapter 4 combines polygenic prediction, a central instrument in personal genomics, with geo-spatial data and regional policy information.

All three chapters push the boundaries of their respective fields by using unusually extensive data sets. With more than 17,000 lost wallets and spanning 40 countries, the experiment in

Chapter 2 is easily among the largest field experiments in behavioral economics. The experiment in Chapter 3 pales in comparison, despite involving a respectable 6,998 Bolsa Família registration centers. This is compensated for by the administrative data, which covers 20 million families for up to six years and enables me to match families extremely precisely to obtain accurate estimates of Bolsa Família's effects in thousands of municipalities. Genetic data are almost proverbially big, and the UK Biobank used in Chapter 4 is no exception, with hundreds of thousands of genetic variants and vast medical histories for approximately 500,000 people. The sheer size of this sample makes it possible to construct a polygenic score with sufficient predictive power for behavioral traits. Likewise, it's the comprehensive list of more than 50,000 pubs and thousands of retail locations that allows me to construct a high-resolution measure of alcohol availability. In short, when it comes to data, big is beautiful.

In conclusion, each of the Chapters offers insight into how our surroundings mitigate or accentuate our inane tendencies—be it for honesty or dishonesty in the case of Chapters 2 and 3, or genetic predisposition in Chapter 4. As new data is unlocked for research and our arsenal of methods for scientific exploration grows, so will our knowledge of human nature and how it is interwoven with the world around us. To belabor the opening analogy one final time, one might compare our understanding of human nature with the Rocky Mountains, as described in the novel *Centennial*:

"The Rockies are therefore very young and should never be thought of as ancient. They are still in the process of building and eroding, and no one today can calculate what they will look like ten million years from now. They have the extravagant beauty of youth, the allure of adolescence, and they are mountains to be loved."

— James A. Michener (2014, p.41)

REFERENCES

- ABELER, J., D. NOSENZO, AND C. RAYMOND (2019): “Preferences for truth-telling,” *Econometrica*, 87, 1115–1153.
- ALGAN, Y. AND P. CAHUC (2013): “Trust and growth,” *Annual Review of Economics*, 5, 521–549.
- ARROW, K. J. (1972): “Gifts and exchanges,” *Philosophy & Public Affairs*, 1, 343–362.
- BECKER, G. S. (1968): “Crime and punishment: an economic approach,” *Journal of Political Economy*, 76, 169–217.
- CHARNESS, G. AND M. RABIN (2002): “Understanding social preferences with simple tests,” *Quarterly Journal of Economics*, 117, 817–869.
- DARWIN, C. (1859): *The Origin of Species; And, the Descent of Man*, Modern library.
- DUVAL, S. AND R. A. WICKLUND (1972): *A theory of objective self awareness.*, Academic press.
- ELLINGSEN, T. AND M. JOHANNESSON (2004): “Promises, threats and fairness,” *Economic Journal*, 114, 397–420.
- FEHR, E. AND K. M. SCHMIDT (1999): “A theory of fairness, competition, and cooperation,” *Quarterly Journal of Economics*, 114, 817–868.
- FISMAN, R., S. KARIV, AND D. MARKOVITS (2007): “Individual preferences for giving,” *American Economic Review*, 97, 1858–1876.
- FREUD, S. (1930): *Civilization and its discontents*, WW Norton & Company.
- GÄCHTER, S. AND J. F. SCHULZ (2016): “Intrinsic honesty and the prevalence of rule violations across societies,” *Nature*, 531, 496–499.
- GNEEZY, U., A. KAJACKAITE, AND J. SOBEL (2018): “Lying aversion and the size of the lie,” *American Economic Review*, 108, 419–53.
- HECKMAN, J. J. (2007): “The economics, technology, and neuroscience of human capability formation.” *Proceedings of the National Academy of Sciences*, 104, 13250–5.
- HENRICH, J., S. J. HEINE, AND A. NORENZAYAN (2010): “Most people are not WEIRD,” *Nature*, 466, 29.
- HUME, D. (1748): *Philosophical essays concerning human understanding*, A. Millar.
- HUNTER, D. J. (2005): “Gene-environment interactions in human diseases.” *Nature Reviews Genetics*, 6, 287–98.
- KRUPKA, E. L. AND R. A. WEBER (2013): “Identifying social norms using coordination games: Why does dictator game sharing vary?” *Journal of the European Economic Association*, 11, 495–524.
- MAZAR, N., O. AMIR, AND D. ARIELY (2008): “The dishonesty of honest people: A theory of self-concept maintenance,” *Journal of Marketing Research*, 45, 633–644.
- MICHENER, J. A. (2014): *Centennial: A Novel*, New York, NY: Dial Press Trade Paperbacks.
- MOKDAD, A. H., J. S. MARKS, D. F. STROUP, AND J. L. GERBERDING (2004): “Actual causes of death in the United States, 2000.” *JAMA*, 291, 1238–45.

MULCASTER, R. (1582): *Mulcaster's Elementaire*, London: Clarendon Press.

ROOM, R., T. BABOR, AND J. REHM (2005): "Alcohol and public health," *The Lancet*, 365, 519–530.

STIGLER, G. J. AND G. S. BECKER (1977): "De gustibus non est disputandum," *American Economic Review*, 67, 76–90.

Chapter 2

Civic Honesty around the Globe*

ALAIN COHN, MICHEL MARÉCHAL,
DAVID TANNENBAUM, AND CHRISTIAN ZÜND

Abstract: Civic honesty is essential to social capital and economic development, but is often in conflict with material self-interest. We examine the trade-off between honesty and self-interest using field experiments in 355 cities spanning 40 countries around the globe. We turned in over 17,000 lost wallets with varying amounts of money at public and private institutions, and measured whether recipients contacted the owner to return the wallets. In virtually all countries citizens were more likely to return wallets that contained more money. Both non-experts and professional economists were unable to predict this result. Additional data suggest our main findings can be explained by a combination of altruistic concerns and an aversion to viewing oneself as a thief, which increase with the material benefits of dishonesty.

Acknowledgments: We are grateful for many helpful discussions including those of Johannes Abeler, Michael Bauer, Aline Bütikofer, Stefano DellaVigna, Ernst Fehr, Ray Fisman, Serra Marta Garcia, Joe Henrich, Magnus Johannesson, Emir Kamenica, Johann Graf Lambsdorff, Vai-Lam Mui, Nick Netzer, Andrew Oswald, Devin Pope, Gautam Rao, Andrei Shleifer, Paul Smeets, Richard Thaler, Fabrizio Zilibotti and audiences at the University of California, San Diego, Harvard University, Monash University, Stanford Institute for Theoretical Economics, European University Institute, Florence, University of Michigan, University of Nottingham, University of Amsterdam, London School of Economics and Political Science, University of Copenhagen, Copenhagen Business School, CERGE-EI Prague, Helsinki Center of Economic Research, University of Chicago, Ohio State University, Laval University, University of Mainz, University of Munich, University of Stockholm, University of Vienna, WZB Berlin, University of Lyon, and University of Zurich. We thank Jan Aeberhard, Marie Baumann, Karim Ben Hassine, Dominic Bigliel, Thomas Braschler, Pascal Bührig, Flavio Caderas, Cosma Gabaglio, Christine Kaut, Vaclav Korb, Flurin Noldin, Nenad Ruvidic, Nicolas Sampl, Bruno Scherrer, Marco Schwarz for outstanding research assistance and Andreas Saurer for excellent technical assistance. We are grateful for financial support from the Gottlieb Duttweiler Institute.

*A version of this chapter has been published as: Cohn, A., Maréchal, M. A., Tannenbaum, D., & Zünd, C. L. (2019). Civic honesty around the globe. *Science*, 365(6448), 70–73.

I. INTRODUCTION

Honest behavior is a central feature of economic and social life (Arrow, 1972; Algan and Cahuc, 2013). Without honesty, promises are broken, contracts go unenforced, taxes remain unpaid, and governments become corrupt. Such breaches of honesty are costly to individuals, organizations and entire societies. For example, losses due to tax evasion in the US are estimated in the hundreds of billions of dollars each year (IRS, 2016), and the global cost of corruption and other illicit financial flows has been estimated at 1.3 trillion dollars annually—an amount roughly equal in size to the gross domestic product of Australia (Kar and Freitas, 2011; Bank, 2017).

In this paper we examine how acts of civic honesty, where people voluntarily refrain from opportunistic behavior, are affected by monetary incentives to act otherwise. Although there is robust experimental literature on the conditions that give rise to honest behavior (Ellingsen and Johannesson, 2004; Mazar et al., 2008; Shalvi et al., 2015; Gächter and Schulz, 2016; Gneezy et al., 2018; Abeler et al., 2019), little is known about how material incentives impact civic honesty, particularly in field settings. Understanding the relationship between civic honesty and material incentives is not only practically relevant, but also theoretically important.

Theories of honesty make different predictions about the role of material incentives. Classic economic models based on rational self-interest suggest that, all else equal, honest behavior will become less common as the material incentives for dishonesty increase (Becker, 1968). Models of human behavior that incorporate altruistic or other-regarding preferences also predict dishonesty to rise with increasing incentives, as self-interest virtually always dominates concerns for the welfare of others—we care about others but not as much as we care about ourselves (Fehr and Schmidt, 1999; Charness and Rabin, 2002; Fisman et al., 2007). As a result, self-interest will play an increasingly prominent role in behavior as the material incentives for dishonesty grow. Psychological models based on self-image maintenance predict that people will cheat for profit so long as their behavior does not require them to negatively update their self-concept (Mazar et al., 2008; Duval and Wicklund, 1972). However, it is unclear *ex ante* whether self-image concerns will become more or less important as the incentives for dishonesty increase, and what form that relationship will take. A further complication is that most of the experimental literature on honest behavior involves modest financial stakes, has been conducted in laboratory settings (where people understand their behavior is being observed), and tends to rely on populations from Western, educated, industrialized, rich and democratic societies (Henrich et al., 2010).

II. FIELD EXPERIMENT

We visited 355 cities in 40 countries and turned in a total of 17,303 wallets. We typically targeted the five to eight largest cities in a country, with roughly 400 observations per country.

Wallets were returned to one of five societal institutions: (i) banks, (ii) theaters, museums, or other cultural establishments, (iii) post offices, (iv) hotels, and (v) police stations, courts of law, or other public offices. These institutions serve as useful benchmarks because they are common across countries and typically have a public reception area where we could perform the drop-offs.

A. Design

Our wallets were transparent business card cases, which we used to ensure that recipients could visually inspect without having to physically open the wallet (Figure A2.1 in Appendix A2.I). Our key independent variable was whether the wallet contained money, which we randomly varied to hold either no money or US \$13.45 (“NoMoney” and “Money” conditions, respectively). We used local currencies, and to ensure comparability across countries, we adjusted the amount according to each country’s purchasing power. Each wallet also contained three identical business cards, a grocery list, and a key. The business cards displayed the owner’s name and email address, and we used fictitious but commonplace male names for each country. Both the grocery list and business cards were written in the country’s local language to signal that the owner was a local resident.

After walking into the building, one of our research assistants (from a pool of eleven male and two female assistants) approached an employee at the counter and said, “Hi, I found this [pointing to the wallet] on the street around the corner.” The wallet was then placed on the counter and pushed over to the employee. “Somebody must have lost it. I’m in a hurry and have to go. Can you please take care of it?” The research assistant then left the building without leaving contact details or asking for a receipt. Our key outcome measure was whether recipients contacted the owner to return the wallet. We created unique email addresses for every wallet and recorded emails that were sent within 100 days of the initial drop-off. Complete methods and results, including additional robustness checks such as testing for experimenter effects, can be found in Appendix A2.

B. Results

As shown in the left half of Figure 2.1, our cross-country experiments return a remarkably consistent result: citizens were overwhelmingly more likely to report lost wallets with money than without. We observed this pattern for 38 out of our 40 countries, and in no country did we find a statistically significant decrease in reporting rates when the wallet contained money. On average, adding money to the wallet increased the likelihood of reporting a wallet from 40% in the NoMoney condition to 51% in the Money condition ($P < 0.0001$). This result holds when controlling for a number of recipient and situational characteristics (Table A2.8 in Appendix A2.III). Furthermore, while rates of civic honesty vary substantially from country to country, the absolute increase in honesty across conditions was stable. As shown on the right

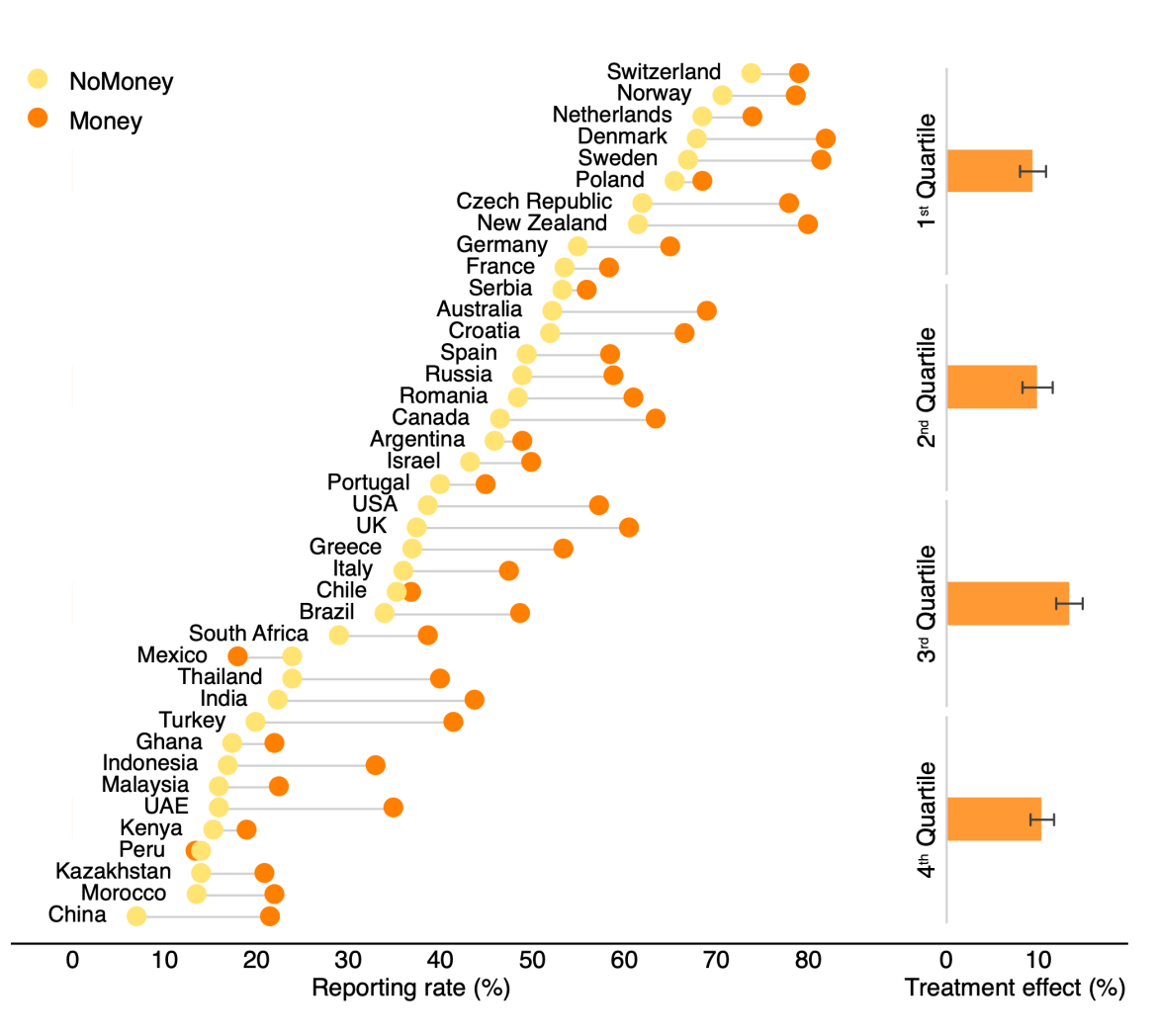


FIGURE 2.1

Share of Wallets Reported in the NoMoney and Money Condition by Country

Notes. Left hand side: Share of wallets reported in treatments NoMoney (US \$0) and Money (US \$13.45) by country. The amount of money in the wallet is adjusted according to each country's purchasing power. Right hand side: Average difference between treatment Money and NoMoney across quartiles based on absolute response rates in the NoMoney condition. Error bars represent standard errors of the mean.

half of Figure 2.1, the average treatment effect is roughly equal in size across quartiles based on absolute response rates.

Citizens displayed greater civic honesty when the wallets contained money, but perhaps this is because the amount was not large enough to be financially meaningful. To examine this possibility we also ran a “BigMoney” condition in three countries (US, UK, and Poland) that increased the money inside the wallet to US \$94.15, or seven times the amount in our original Money condition. Shown in Figure 2.2, reporting rates in all three countries increase even further when the wallets contained a sizable amount of money. Pooled across the three

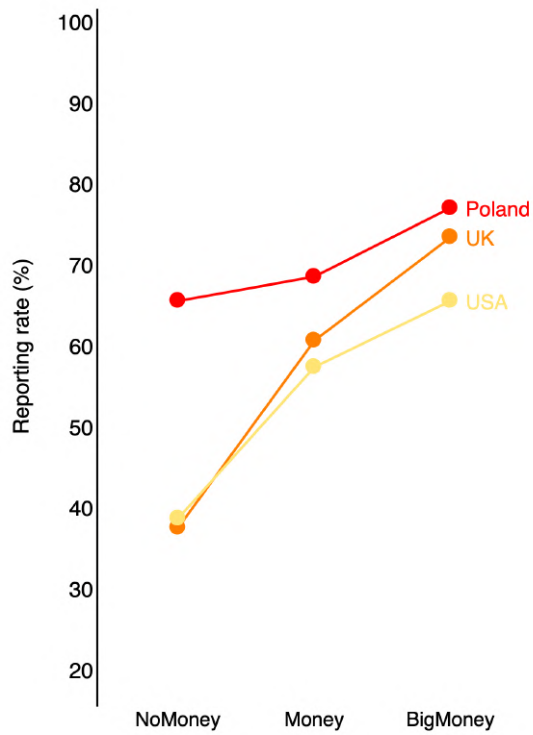


FIGURE 2.2
Reporting Rates as a Function of Monetary Stakes
Notes. Share of wallets reported in the NoMoney (US \$0), Money (US \$13.45), and BigMoney (US \$94.15) conditions.

countries, response rates increased from 46% in the NoMoney condition to 61% in the Money condition, and topped out at 72% in the BigMoney condition ($P < 0.0001$ for all pairwise comparisons; Table A2.9 in Appendix A2.III).

III. MECHANISM

We next turn to the question of why people are especially likely to return a lost wallet when it contains more, rather than less, money. Our study design allows us to rule out several possible explanations. We first explored the possibility that recipients were worried about legal penalties for failing to return a wallet, especially when the wallet contained increasing amounts of money. To address this issue, we examined whether relative reporting rates were affected by (a) the presence of other individuals when receiving the wallet, (b) the presence of security cameras in the building, and (c) state-level variation in lost property laws within the United States. Civic honesty should increase as a function of these variables if recipients were concerned about possible punishment or probability of detection, yet we find that none of these factors explain meaningful variation in reporting rates across treatment conditions (Tables A2.14-A2.16

in Appendix A2.IV). A second explanation is that since we only measured whether recipients reported a lost wallet, recipients in the money conditions may have been more likely to return the wallets while pocketing the cash. We conducted an audit on a subset of wallets reported to us and do not find support for this explanation: over 98% of the money in the wallets we collected was returned. A third possible explanation is that recipients expected a larger “finder’s fee” upon returning wallets with greater amounts of money. Using national representative surveys conducted in the US, UK, and Poland, we asked respondents the size of the reward they would expect upon returning a wallet with the amounts of money we used in our studies. We fail to find evidence that people expect a larger reward for returning a wallet with more, rather than less, money (Table A2.17 in Appendix A2.IV).

Having ruled out three possible explanations, we next formulate and test a simple behavioral model that captures the pattern of results observed in the data (full model details can be found in Appendix A2.II). In our framework, civic honesty is determined by the interplay between four components: (i) the economic payoff of keeping the wallet, (ii) the fixed effort cost of contacting the wallet’s owner, (iii) an altruistic concern for the owner’s welfare, and (iv) the costs associated with negatively updating one’s self-image as a thief (what we will call “theft aversion”).

A key feature of our framework is that altruistic concerns are affected by the contents of the wallet thought to be valuable to the owner, whereas concerns of theft aversion are only affected by the contents of the wallet that are also valuable to the recipient (e.g., money). To distinguish between these two motivations, we conducted a “Money-NoKey” condition in our US, UK, and Poland locations with wallets identical to our Money condition but which did not contain a key. Unlike money, the key is valuable to the owner but not to the recipient, and so any difference between the Money and Money-NoKey conditions can be ascribed to altruistic concerns. Shown in Table A2.10 in Appendix A2.III, recipients were on average 9.2 percentage points more likely to report a wallet with a key than without ($P = 0.0001$ when results are pooled across countries). This suggests that recipients reported a lost wallet partly because recipients are concerned about the harm they impose on the owner.

The second part of our framework—and crucial to explaining the increase in reporting rates for wallets with greater amounts of money—involves the aversion to viewing oneself as a thief. Using nationally representative surveys conducted in the US, UK, and Poland, we asked respondents to imagine receiving a wallet with the contents in our four conditions (NoMoney, Money, BigMoney, and Money-NoKey) and rated the extent to which failing to return the wallet would feel like stealing on a scale from 0 (*not at all*) to 10 (*very much*). Respondents reported that failing to return a wallet would feel more like stealing when the wallet contained a modest amount of money than when it contained no money, and that such behavior would feel even more like stealing when the wallet contained a substantial amount of money ($P \leq 0.007$ for all pairwise comparisons; Table A2.11 in Appendix A2.III). This tells us that, consistent with our behavioral data on wallet reporting rates, the self-image cost of failing to return the wallet

likely increases with the amount of money in the wallet. By contrast, we fail to observe a reliable difference in “feels like stealing” scores when comparing wallets that contained the same amount of money but differed in whether they also contained a key (Money vs. Money-NoKey; $P = 0.259$). This tells us that concerns of theft aversion are likely tied to contents valuable to the recipient, such as the amount of money inside the wallet, but not to other contents that are only valuable to the owner. Although survey responses do not always generalize to real behavior and should be interpreted carefully, these findings are consistent with the hypothesis that larger monetary payoffs for dishonesty are also associated with increased psychological costs, and that the increase in psychological costs can outweigh the marginal economic benefits of dishonesty.

IV. PREDICTION EXPERIMENTS

In a final set of studies, we investigated whether people anticipate this form of civic honesty. We asked a sample of 299 participants to predict reporting rates in the US for wallets that contained \$0, \$13.45, and \$94.15 (corresponding to our NoMoney, Money, and BigMoney conditions). To encourage accuracy, we notified respondents that the most accurate predictors would receive a cash bonus. Shown in Figure 2.3 (middle), we find that respondents’ beliefs were at odds with the behavioral data (Figure 2.3, left). Respondents predicted that rates of civic honesty would be highest when the wallet contained no money ($M = 73\%$, $SD = 29$), lower when the wallet contained a modest amount of money ($M = 65\%$, $SD = 24$), and lower still when the wallet contained a substantial amount of money ($M = 55\%$, $SD = 29$). The average predicted change in reporting rates from condition to condition was significantly different from the actual change in reporting rates ($P < 0.001$ for all pairwise comparisons). As wallet amounts increased, 64% of respondents incorrectly predicted reporting rates would decrease and 18% correctly predicted reporting rates would increase ($P < 0.001$ by a sign test). Additional questioning suggests that respondents’ predictions reflected a mental model of human behavior that exaggerates the role of narrow self-interest (Kruger and Gilovich, 1999; Miller and Ratner, 1998). When wallets contained more money, respondents expected self-interest to grow and altruistic concerns for the owner to fade, and gave little weight to theft aversion in influencing reporting rates (Table A2.13 in Appendix A2.III).

The general public incorrectly predicts how citizens will respond as the monetary value of the wallet increases, but perhaps professional economists will be more accurate. We asked a sample of 279 top-performing academic economists to predict our results. Like our non-experts, this sample also did not expect reporting rates to increase for wallets with greater amounts of money. Shown in Figure 2.3 (right), respondents on average predicted that rates of civic honesty would be higher in the NoMoney and Money conditions ($M = 69\%$, $SD = 25$ and $M = 69\%$, $SD = 21$, respectively) than in the BigMoney condition ($M = 66\%$, $SD = 23$). These predictions were again significantly different from the actual changes we observe across conditions ($P < 0.001$

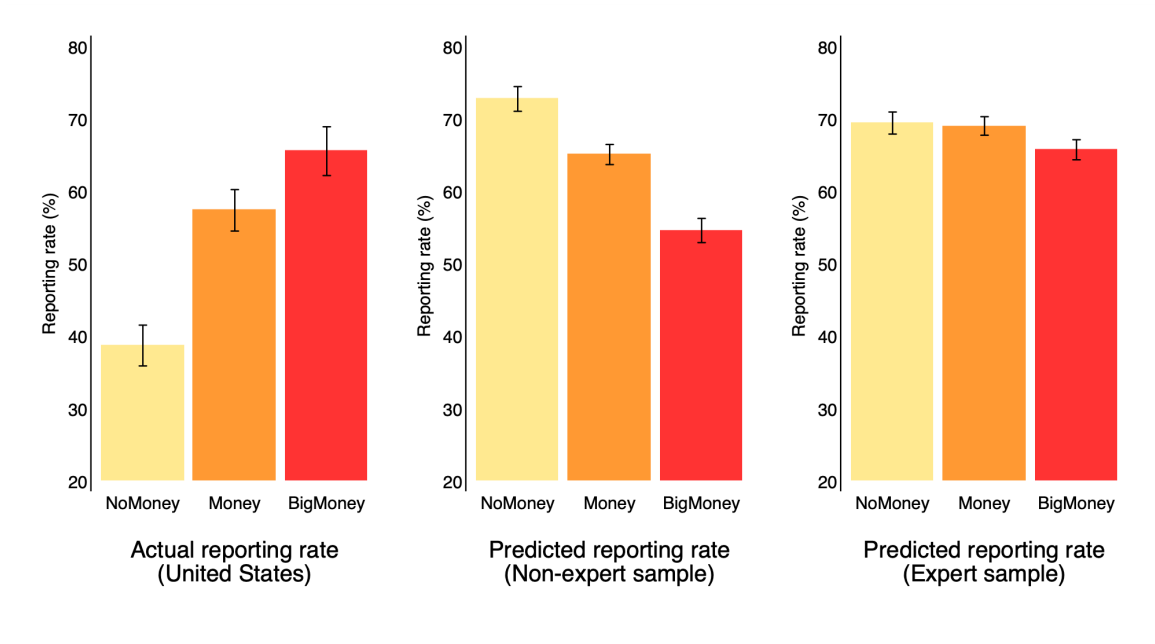


FIGURE 2.3
Actual versus Predicted Reporting Rates

Notes. (left) Actual reporting rates in the US for each condition ($N = 800$). Error bars represent robust standard errors. (middle) Average predicted reporting rates for the US by our non-expert sample ($N = 299$). Error bars represent robust standard errors clustered by participants. (right) Average predicted reporting rates for the US by our expert sample of academic economists ($N = 279$). Error bars represent robust standard errors clustered by participants.

for all pairwise comparisons). However, the degree of miscalibration among economists was less severe than in our non-expert sample. As wallet amounts increased, 49% of economists incorrectly predicted reporting rates would decrease and 29% correctly predicted reporting rates would increase ($P < 0.001$ by a sign test).

V. CONCLUSION

We conducted field experiments in 40 countries to examine whether people act more dishonestly when they have a greater economic incentive to do so, and found the opposite to be true. Citizens were more likely to return wallets that contained relatively larger amounts of money. This finding is robust across countries and institutions, and holds even when economic incentives for dishonesty are substantial. Our results are consistent with theoretical models that incorporate altruism and self-image concerns, but also suggest modification in that non-pecuniary motivations directly interact with the material benefits gained from dishonest behavior. When people stand to heavily profit from engaging in dishonest behavior, the desire to cheat increases but so do the psychological costs of viewing oneself as a thief—and sometimes the latter will dominate the former.

Our findings also represent a unique data set for examining cross-country differences in civic

honesty. Honesty is a key component of social capital (Guiso et al., 2011), and here we provide an objective measure to supplement the large body of work that has traditionally examined social capital using subjective survey measures (Glaeser et al., 2000; Nannestad, 2008; Algan and Cahuc, 2013; Alesina and Giuliano, 2015). Using average response rates across countries, we find substantial variation in rates of civic honesty, ranging from 14% to 76%. This variation largely persists even when controlling for a country's gross domestic product, suggesting that other factors besides country wealth are also at play. In Appendix A2.III, we provide an analysis suggesting that economically favorable geographic conditions, inclusive political institutions, national education, and cultural values that emphasize moral norms extending beyond one's in-group are also positively associated with rates of civic honesty. Future research is needed to identify how these and other factors may contribute to societal differences in honest behavior.

REFERENCES

- ABELER, J., D. NOSENZO, AND C. RAYMOND (2019): "Preferences for truth-telling," *Econometrica*, 87, 1115–1153.
- ALESINA, A. AND P. GIULIANO (2015): "Culture and institutions," *Journal of Economic Literature*, 53, 898–944.
- ALGAN, Y. AND P. CAHUC (2013): "Trust and growth," *Annual Review of Economics*, 5, 521–549.
- ARROW, K. J. (1972): "Gifts and exchanges," *Philosophy & Public Affairs*, 1, 343–362.
- BANK, W. (2017): *World Development Indicators*, The World Bank.
- BECKER, G. S. (1968): "Crime and punishment: an economic approach," *Journal of Political Economy*, 76, 169–217.
- CHARNESS, G. AND M. RABIN (2002): "Understanding social preferences with simple tests," *Quarterly Journal of Economics*, 117, 817–869.
- DUVAL, S. AND R. A. WICKLUND (1972): *A theory of objective self awareness.*, Academic press.
- ELLINGSEN, T. AND M. JOHANNESSON (2004): "Promises, threats and fairness," *Economic Journal*, 114, 397–420.
- FEHR, E. AND K. M. SCHMIDT (1999): "A theory of fairness, competition, and cooperation," *Quarterly Journal of Economics*, 114, 817–868.
- FISMAN, R., S. KARIV, AND D. MARKOVITS (2007): "Individual preferences for giving," *American Economic Review*, 97, 1858–1876.
- GÄCHTER, S. AND J. F. SCHULZ (2016): "Intrinsic honesty and the prevalence of rule violations across societies," *Nature*, 531, 496–499.
- GLAESER, E. L., D. I. LAIBSON, J. A. SCHEINKMAN, AND C. L. SOUTTER (2000): "Measuring trust," *Quarterly Journal of Economics*, 115, 811–846.
- GNEEZY, U., A. KAJACKAITE, AND J. SOBEL (2018): "Lying aversion and the size of the lie," *American Economic Review*, 108, 419–53.

- GUIO, L., P. SAPIENZA, AND L. ZINGALES (2011): “Civic capital as the missing link,” in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. O. Jackson, Amsterdam: North-Holland, vol. 1, 417–480.
- HENRICH, J., S. J. HEINE, AND A. NORENZAYAN (2010): “Most people are not WEIRD,” *Nature*, 466, 29.
- IRS (2016): *Tax Gap Estimates for Tax Years 2008-2010*, Internal Revenue Services.
- KAR, D. AND S. FREITAS (2011): *Illicit Financial Flows from Developing Countries over the Decade Ending 2009*, Global Financial Integrity.
- KRUGER, J. AND T. GILOVICH (1999): ““Naive cynicism” in everyday theories of responsibility assessment: On biased assumptions of bias,” *Journal of Personality and Social Psychology*, 76, 743–753.
- MAZAR, N., O. AMIR, AND D. ARIELY (2008): “The dishonesty of honest people: A theory of self-concept maintenance,” *Journal of Marketing Research*, 45, 633–644.
- MERRITT, C. B. AND R. G. FOWLER (1948): “The pecuniary honesty of the public at large,” *Journal of Abnormal and Social Psychology*, 43, 90–93.
- MILGRAM, S., L. MANN, AND S. HARTER (1965): “The lost-letter technique: A tool of social research,” *Public Opinion Quarterly*, 29, 437–438.
- MILLER, D. T. AND R. K. RATNER (1998): “The disparity between the actual and assumed power of self-interest,” *Journal of Personality and Social Psychology*, 74, 53–62.
- NANNESTAD, P. (2008): “What have we learned about generalized trust, if anything?” *Annu. Rev. Polit. Sci.*, 11, 413–436.
- SHALVI, S., F. GINO, R. BARKAN, AND S. AYAL (2015): “Self-serving justifications: Doing wrong and feeling moral,” *Current Directions in Psychological Science*, 24, 125–130.

Chapter 3

Local Corruption, Income Underreporting, and Policy Effectiveness

CHRISTIAN ZÜND

Abstract: This paper demonstrates how corruption negatively impacts policy effectiveness in a program that is explicitly designed to be unsusceptible to bribery, clientelism, and the embezzlement of money: Brazil's Bolsa Família program. The program's centralized beneficiary selection process enables me to identify families that were randomly admitted or not admitted to Bolsa Família and to estimate the program's effectiveness in different years and municipalities. Exploiting a second natural experiment, I then show that Bolsa Família's effect on school enrollment increases by a third after a municipality has been audited at random. Using a theoretical model, administrative data, and a field experiment with 6,998 registration centers, I find that local corruption increases the probability that families successfully underreport their income when registering for Bolsa Família, making it harder to target the families that benefit most. Ruling out alternative explanations, I show that neither changes in social norms, improved school attendance monitoring, administrative processes, infrastructure, complementary programs, tighter governance, nor increased whistleblowing can account for the increase in Bolsa Família's effectiveness after a municipality has been audited at random. Taken together, these results suggest that income underreporting can explain how, despite the program's safeguards, local corruption undermines the effectiveness of Bolsa Família.

Acknowledgments: I am grateful for many helpful discussions including those with Johannes Abeler, Silvia Barcellos, Pietro Biroli, Leandro Carvalho, Alain Cohn, Raymond Duch, Charles Efferson, Flávio Éiro, Ernst Fehr, Michel Maréchal, Sonja Vogt, and audiences at the University of Oxford and University of Zurich. I thank Fares Alkudmani, Paula Barcelos, Danilo Goncalves, Patricia Vasconcellos, Felix Wüthrich, Ninjas Zegg, and Corina Zünd for outstanding research assistance. I am grateful for financial support from the Swiss National Science Fund.

I. INTRODUCTION

Corruption is costly. While the direct costs of corruption—money diverted, embezzled, stolen—are substantial, they may well be outweighed by indirect effects (Kaufmann, 2005) such as underinvestment in human capital (Mo, 2001), higher child mortality (Gupta et al., 2001), decreased private investment (Mauro, 1995), lower returns to public investment (Tanzi and Davoodi, 1998), and slower economic growth (Mauro, 1995, 2004). This literature has mostly focused on two consequences of corruption: the effect of missing funds (through embezzlement, graft, inappropriate procurement) and of preferential access to public goods and services (in exchange for bribes, votes, or because of family or group membership). Leakage of funds and preferential access to public goods reduces the effectiveness of governments to provide adequate education (e.g., Reinikka and Svensson, 2005; Ferraz et al., 2012; Abdulai and Hickey, 2016), run medical systems (e.g., McPake et al., 1999; Holmberg and Rothstein, 2011; Mostert et al., 2015), enforce traffic regulation (e.g., Bertrand et al., 2007; Olken and Barron, 2009), and operate poverty-relief programs (e.g., Olken, 2006; Penfold-Becerra, 2007). However, even if effective safeguards against bribery, clientelism, and embezzlement are in place, government programs are not necessarily immune to the corrosive impact of corruption.

This paper shows how corruption negatively impacts policy effectiveness in a program that is explicitly designed to be unsusceptible to bribery, clientelism, and the embezzlement of money. Brazil’s Bolsa Família promotes educational participation by paying a monthly benefit to approximately 14 million families if their children attend school regularly. These funds are transferred directly to the beneficiaries so that local officials cannot pocket the money. Not only are local officials bypassed in the payment process, they are also not involved in selecting the families that benefit from Bolsa Família; the selection of beneficiaries is anonymized and conducted through a central process to prevent local officials from controlling access to the program to extract rents or to benefit their supporters. As a result, Bolsa Família is generally considered an exception to Brazil’s otherwise widespread clientelism (Sugiyama and Hunter, 2013).¹

One possible explanation for the negative effect of corruption is that it enables families to underreport their income to gain or keep Bolsa Família; even if corrupt *officials* can not embezzle funds, local corruption can make it easier for *families* to cheat. Accusations of families underreporting their income are common both in the media and in the political discussion (e.g., Brazil, 2016; Caram, 2016; Fausto Macedo, 2016; O Globo, 2016). However, although it is generally assumed that income underreporting decreases the effectiveness of programs like Bolsa Família, this assumption has not previously been tested. This paper provides causal evidence

1. The prevention of clientelism at the local level does not ensure that politicians at the federal level cannot expand the program for electoral reasons. How large these electoral gains are and whether this constitutes a form of clientelism is hotly debated (e.g., Hunter and Power, 2007; Zucco, 2009; Bohn, 2011; Daëff, 2015) but is beyond the scope of this paper.

linking local corruption, families' ability to successfully underreport their income, and lower causal effects of Bolsa Família on school enrollment.

Estimating the impact of corruption on the effectiveness of government policy is challenging for two main reasons: First, it requires that corruption varies sufficiently much between the places or during the time where the programs are evaluated—with all the usual caveats about the endogeneity of corruption. To address this challenge, I use a well-established natural experiment, Brazil's audit lottery program. These audits of randomly selected municipalities have been shown to decrease corruption in subsequent years (Avis et al., 2018), thus providing exogenous variation in corruption between different places and years.

The main challenge is to measure the causal effect of the policy on school enrollment in different regions and times, so that changes in its effectiveness can be estimated. While it is common to use the accuracy of targeting as a proxy for the effectiveness of anti-poverty programs, this is primarily done because no appropriate measure of program effectiveness is available (De Janvry et al., 2012). Indeed, as pointed out by Ravallion (2009), better targeting does not guarantee that a program is actually more effective. Thus, even if there is plausibly exogenous variation in corruption, one also needs a second identification strategy that provides causal estimates of the effects of the policy in different regions or times.

To estimate the effectiveness of Bolsa Família in different regions and times, I develop a novel identification strategy that makes use of the program's central automated selection process. By using data from the Cadastro Único, the official database used for beneficiary selection, I can reconstruct the algorithm's priority strata to identify families on the margin of the program that were randomly included or not included in Bolsa Família. Because the beneficiary selection process uses only data in the Cadastro Único, there are no *unobservable* differences that affect who gets included in the program. This built-in conditional independence allows me to estimate the causal effect of Bolsa Família on school enrollment for different years, priority strata, and municipalities. The identification strategy passes several validation tests, and it successfully recovers the positive impact of the program documented by others (e.g., Cardoso and Souza, 2003; Glewwe and Kassouf, 2012; Schaffland, 2012; De Brauw et al., 2015).

This identification strategy has several advantages over existing approaches to estimating the effect of Bolsa Família. First, because it uses the same data as the selection algorithm, the identification strategy can identify families that were actually on the margin of the program, unlike many evaluations that rely on propensity scores to match families with similar probabilities of being included in Bolsa Família (e.g., Cardoso and Souza, 2003; Schaffland, 2012). Second, the estimated treatment effects are unbiased, conditional on self-reported income, because the identification strategy uses the official database to identify marginal families based on their income *before* they are included in Bolsa Família.² Third, unlike approaches that use variation in

2. Self-reported incomes in other data sets are likely to differ from data in the Cadastro Único, where families have stronger incentives to underreport their incomes. Moreover, relative to non-beneficiaries, beneficiary

the program's rollout or geographic coverage (e.g., Glewwe and Kassouf, 2012), it can estimate the effectiveness of the program at the municipality level and at different times. Finally, while De Janvry et al. (2012) measure policy effectiveness at the municipality level, they estimate it for Bolsa Escola, Bolsa Família's predecessor program. One crucial difference between the two programs is that municipalities selected the beneficiaries for Bolsa Escola. As a result, the critical assumption that future beneficiaries and non-beneficiaries have similar trends in educational participation, conditional on child characteristics, need not hold for Bolsa Escola,³ but it is guaranteed by Bolsa Família's centralized beneficiary selection process.

The first main result of the paper is that Bolsa Família is less effective in municipalities with more corruption, even though local officials cannot divert money from the program and have no say in the selection of beneficiaries. Using Brazil's initiative to randomly audit municipalities as an exogenous shock to corruption, I show that Bolsa Família's effect on school enrollment increases by about a third after a random audit. This finding is highly robust. Most importantly, it persists if the sample is defined more restrictively and for different specifications of the two natural experiments.

I then test whether income underreporting can account for this result. Because municipalities are responsible for registering potential beneficiaries, local corruption can play the most prominent role at the registration stage. Using a theoretical model, administrative data, and a field experiment, I show that local corruption increases the probability that families successfully underreport their income. This makes it harder to target the families that benefit most from Bolsa Família, which decreases the effectiveness of the program.

I consider a simplified model of Bolsa Família's registration process, where families decide what income to report, and the families with the lowest self-reported incomes are included in the program. If a family's reported income deviates too much from its actual income, it risks being caught and the probability of detection is lower in municipalities with more corruption. The first part of the registration model predicts that after a random audit income underreporting decreases together with local corruption. As more families underreport their income, Bolsa Família can no longer target the families that benefit most, and its effectiveness decreases. Thus, the second part of the model explains how Bolsa Família becomes more effective after a random audit and makes testable predictions about the profile of families that will see the most substantial effectiveness gains.

households are more likely to underreport their incomes in other surveys in order to avoid detection, so that methods that match or weight beneficiary and non-beneficiary families based on their reported income *after* treatment assignment (e.g., Cardoso and Souza, 2003; De Brauw et al., 2015) are potentially biased. Strategic income reporting is also problematic for regression discontinuity designs (e.g., Schaffland, 2012) that rely on the assumption that families with a self-reported income around Bolsa Família's income threshold are relatively similar.

3. Indeed, as program effectiveness depends on selecting the children with the biggest expected *gains* from inclusion, municipalities' selections might well be affected by characteristics that are unobservable to the researchers.

As predicted by the registration model, fewer families claim to have zero income, and fewer families report an income that is eligible for Bolsa Família after a municipality has been audited at random. Moreover, just after a random audit the number of ineligible families that are detected and excluded from the Bolsa Família program skyrockets, but then falls to a lower level than before the audit, suggesting that the audits indeed increase the probability that ineligible families are detected and discourage new registrants from underreporting their income. As the second part of the model predicts, the treatment effect increases most for families that were not visited by a social worker during the registration process and for families that have the strongest incentives to misrepresent their incomes.

As direct evidence for the income underreporting explanation, I conducted a field experiment with 6,998 Bolsa Família registration centers to show that local administrators are indeed less likely to register ineligible families after a random audit. Fictitious applicants asked about the possibility of receiving Bolsa Família, and their characteristics were experimentally varied to make them eligible or ineligible while holding everything else constant. Consistent with the income underreporting mechanism, registration centers in audited municipalities are significantly less likely to engage with ineligible families and to incorrectly state that a sender's income is compatible with Bolsa Família. Taken together, these findings suggest that the effectiveness gains of Bolsa Família are the result of improvements in the way the program is targeted to the families that benefit from it the most.

The lower rates of income underreporting after a random audit might also reflect changes in social norms, rather than differences in the difficulty of successfully misrepresenting one's income. For example, the experience of being audited and the revelation of the irregularities in Bolsa Família might change citizens' social norms about underreporting their income, condoning public corruption, or reporting suspected fraud. To test this alternative hypothesis, I elicited relevant social norms in an incentivized online experiment with 675 participants living in 424 municipalities, some of which had been randomly selected for audits in the past. Using the Krupka and Weber (2013) norm elicitation task, I find no evidence that social norms change as a result of a random audit.

Finally, I test several alternative explanations for the increased effectiveness of Bolsa Família after a random audit. To show that embezzled or otherwise diverted funds can indeed not explain the effect of local corruption on the program's effectiveness, I quantify the maximum amount of funding that local administrators can pocket through various strategies. Less than 0.0001% of funds are paid to stolen benefit cards, and the rate of income underreporting is more than ten times smaller among public servants than in the general population. Thus, the financial damages are negligible and cannot account for the effectiveness gains. As municipalities are also responsible for monitoring school attendance, I test whether improvements in monitoring contribute to the effect. Moreover, data from Bolsa Família's internal control programs show that unintentional errors and outdated information do not account for the gains in program

effectiveness. Similarly, neither changes in administrative processes, nor better-equipped registration centers can explain the results. Finally, I rule out that improved oversight is behind the effectiveness gains, using data on municipalities' social governance councils and Bolsa Família's whistleblowing systems.

To my knowledge, this paper presents the first evidence that links income underreporting to lower treatment effects of Bolsa Família or a similar conditional cash transfer program. While evidence for income underreporting is not new,⁴ showing that it can account for the negative impact of corruption on the program's effectiveness fills a significant gap in the literature given the attention it receives in the public discussion of Bolsa Família and similar conditional cash transfer programs. The results suggest that efforts to detect underreporting by Brazil's federal government (e.g., Brazil, 2016) probably increase the effectiveness of Bolsa Família in the long run.

The previous research that the findings of this paper are most closely related to is De Janvry et al. (2012), who study the effects of electoral incentives on the effectiveness of Bolsa Escola and show that the program is more effective at reducing school dropout if a mayor faces reelection. My paper differs from theirs in three essential aspects: First, the focus of my paper is on the causal impact of corruption, whereas De Janvry et al. (2012) study electoral incentives. Second, municipalities were in charge of selecting the beneficiaries under Bolsa Escola. Thus, my paper shows that even with Bolsa Família's additional safeguards against corruption, program effectiveness is still affected by local corruption. Finally, as mentioned earlier, Bolsa Família's centralized selection process allows me to make a much stronger case against selection and unobserved variable bias when estimating the program's effectiveness. The results are also related to a recent paper by Brollo et al. (2019), who provide evidence that mayors strategically manipulate the enforcement of school attendance conditionalities for electoral reasons. While I find no evidence that school attendance monitoring contributes substantially to the gains in program effectiveness, the results reported by Brollo et al. (2019) suggest that electoral motives might explain why some administrators are more lenient when verifying families' incomes.

This paper contributes to the discussion of the effective targeting of social programs. The discussion commonly focuses on the advantages and drawbacks of different targeting schemes (e.g., Alderman, 2002; Ravallion, 2008; Alatas et al., 2012; Stoeffler et al., 2016) or the financial consequences of mistargeting and elite capture (Alatas et al., 2019). However, better targeting does not necessarily imply that anti-poverty programs also perform better (Ravallion, 2009), and the actual implications of mistargeting for the effectiveness of these programs are rarely tested. The evidence presented here makes a strong case that improved targeting indeed increases the

4. In addition to cases uncovered by the press and by government audits, there is also the alarming observation that the number of families receiving Bolsa Família exceeds the number of eligible families estimated from the census in a significant share of municipalities (Fried, 2012). Moreover, Firpo et al. (2014) describe irregularities in the distribution of reported incomes around the eligibility threshold that are consistent with income underreporting or with a temporary reduction of labor supply to qualify for the program.

effectiveness of Bolsa Família. Given the importance of accurate self-reporting and the challenges of income verification for the effective targeting of Bolsa Família and similar programs, the final section of this paper suggests possible interventions to increase the likelihood of accurate reporting.

Taken together, the results of this paper suggest that income underreporting can explain how—despite Bolsa Família’s safeguards against bribery, clientelism, and embezzlement—local corruption undermines the effectiveness of the program. Even though Bolsa Família is often cited as an example of how to safeguard anti-poverty programs against corruption, it is significantly more effective after a random anti-corruption audit. Thus, the results highlight the positive effect of government audits (e.g., Di Tella and Schargrotsky, 2003; Olken, 2007; Bobonis et al., 2016; Avis et al., 2018).

The paper is organized as follows: Section II describes the relevant institutional details of the Bolsa Família program. Section III discusses how the beneficiary selection process can be used to estimate the causal effects of the program. Section IV shows the main finding of the paper, that local corruption decreases the effectiveness of Bolsa Família. Section V introduces a theoretical model of how income underreporting affects the program’s effectiveness, which is then tested using administrative data and a field experiment. Section VI addresses the limited ways in which corrupt local officials can benefit financially from Bolsa Família. Section VII examines to what degree other explanations can contribute to the effect. Section VIII discusses the policy implications of the paper and concludes.

II. INSTITUTIONAL BACKGROUND: BOLSA FAMÍLIA PROGRAM

Brazil’s Bolsa Família program bypasses local officials for both the payment and the beneficiary selection process, making it a suitable setting to study effects of corruption other than clientelism and embezzlement. This section describes four features of the Bolsa Família program that are relevant for this study: First, I explain how Bolsa Família incentivizes educational participation through conditional cash transfers, which allows me to use the program’s impact on school enrollment as a measure of its effectiveness. Second, I turn to the centralized process to select the families to include in Bolsa Família, which forms the backbone of the identification strategy that allows me to identify the program’s effectiveness in different municipalities. I then discuss Bolsa Família’s safeguards against embezzlement and abuse of the program for electoral gains, before turning to its remaining vulnerabilities to other forms of corruption.

A. *Bolsa Família’s Incentives for Educational Participation*

Bolsa Família, the world’s largest conditional cash transfer program, covers approximately 14 million families and pays a monthly benefit provided that families comply with several conditionalities, including regular school attendance. It is arguably Brazil’s main federal initiative

to increase educational participation, both in terms of its size and its prominence in the public discourse.

In addition to guaranteeing a minimum income for extremely poor households, Bolsa Família seeks to combat the intergenerational transmission of poverty by conditioning some of the payments on school attendance, vaccination, and medical checkups (Lindert et al., 2007). To this aim, the program combines unconditional benefits to families in extreme poverty with conditional transfers to poor households with pregnant women and children.⁵ The program was created in early 2004 through a merger of four predecessor programs; Fome Zero and Bolsa Alimentação were focused on nutrition, Auxílio Gas subsidized cooking gas, and Bolsa Escola was a conditional cash transfer program to increase school enrollment.⁶ The consolidation of Brazil's anti-poverty initiatives began with the creation of a shared centralized database,⁷ the Cadastro Único, in 2001 and culminated with the establishment of the Ministério do Desenvolvimento Social (MDS) in 2004. The Cadastro Único serves as the main registry for data on potential beneficiaries of Brazil's anti-poverty programs and is maintained by the MDS and the state-owned federal savings bank, the Caixa Econômica Federal (Caixa).

Under the original rules, households with a per capita income of less than R\$ 50 (the extreme poverty line) were considered extremely poor and received an unconditional basic transfer of R\$ 50, and all families with a per capita income of less than R\$ 100 (the poverty line) were eligible for variable benefits of R\$ 15 for up to three pregnant women or children aged 0 to 15, provided that they comply with the educational and health conditionalities. Although most closely associated with the government of Luiz Inácio Lula da Silva, every government since has significantly expanded eligibility and benefits under the program: The eligibility thresholds have subsequently increased to R\$ 85 and R\$ 170, the basic benefit has been increased to R\$ 89, variable benefits stand at R\$ 41 are now paid for up to five children ages 0 to 15 and one pregnant female, and additional benefits for adolescents (16 to 17) were introduced in 2012 (currently paying R\$ 48 for up to two adolescents). Also introduced was an additional payment individually calibrated such that beneficiaries' post-transfer income per capita (including Bolsa Família) reaches the extreme poverty threshold.

Once a family is included in Bolsa Família, it is required to adhere to a schedule of medical checkups and vaccinations for pregnant women, nursing mothers, and young children, and to ensure that children aged 6 to 15 attend school at least 85% of the time and adolescents at least 75% of the time. The exact conditionalities are decided by the ministries of health and education, that then train municipal workers to monitor and report program compliance through the federal government's online systems. Failure to comply initially results in a warning. If during the next

5. For the purpose Bolsa Família, Lei n° 10.836 defines a family as a "nuclear family, possibly extended by other individuals who have kinship or affinity with it, who form a domestic group, live under the same roof, and support each other." I will use the terms family and household interchangeably.

6. Lei n° 10.836 (January 9, 2004)

7. Decreto n° 3.877 (July 24, 2001)

six months the family does not fulfill the conditionalities, the payment is withheld for one month but can be withdrawn in the next if the family addresses the problem. A third or fourth month of non-compliance results in a two-month suspension of the benefit. During this time no benefit is paid and the family is on probation. Failure to comply during this period results in the cancellation of the benefit and exclusion of the family from the program (MDS and SENARC, 2018).

The federal government can also impose sanctions if a family does not regularly update its data, does not withdraw money for several months, reaches a per capita income that exceeds half a minimum wage, or if it detects that a family has provided false information during the registration process (MDS and SENARC, 2015).

B. Bolsa Família's Beneficiary Selection Process

Partly as a result of Bolsa Família's expansion, coverage of the program is not always universal. Thus, for the time of the analysis (2012-2017), the official database, the Cadastro Único, contains data on both beneficiary families and eligible non-beneficiary families. The number of available places in Bolsa Família depends on the federal funding provided for the program and the municipality's official poverty rate, which is only periodically updated. The program's beneficiary selection process guarantees that only family characteristics that are observable in the Cadastro Único affect who is included in the Bolsa Família program. In Section III, I describe in detail how this process can be used to estimate the causal effects of Bolsa Família by constructing otherwise identical treatment and control groups.

The allocation of benefits is split into four phases:

1. Registration: Families with an income of less than half a minimum wage (currently R\$ 499) register in the Cadastro Único, either at the Centro de Referência de Assistência Social (CRAS), the local center for social assistance, or with a social worker during a home visit. This step happens at the municipal level. A social worker then inputs the data into the Cadastro Único, which is maintained by the Caixa Econômica Federal (Caixa). Families have to update their data at least every other year.
2. Qualification: Once a month, Caixa generates aggregated reports for each municipality with the number of qualifying families in different vulnerability categories. It first extracts data on all qualifying families—families with an income below the current threshold for Bolsa Família that have complete and updated information and are not excluded from the program because of sanctions.⁸ These families are categorized by vulnerability (e.g., indigenous families, families with suspected child labor, families that benefit from another

8. Families can be excluded from Bolsa Família for at least a year for severe infractions, such as underreporting their income. More on this in Section V.

social program)⁹ and the Caixa then produces aggregated reports to determine the number of benefits allocated to each vulnerability category in each municipality.

3. Selection: Based on these aggregate reports, the Secretaria Nacional de Renda de Cidadania (SENARC) decides how many benefits it allocates to each vulnerability category and each municipality. SENARC uses an algorithm that optimizes the allocation of benefits to each category, subject to a budget constraint and additional parameters that allow the SENARC to prioritize especially vulnerable categories.¹⁰ Once it has decided how many benefits it allocates to families in each category, SENARC distributes the benefits across municipalities and prioritizes places with a lower Bolsa Família coverage rate relative to the official municipal poverty rates from the census. SENARC then instructs the Caixa how many benefits to grant to families in each municipality and category.
4. Concession: In the final phase, the Caixa's computer system determines the actual beneficiaries in each vulnerability category and municipality based on families' per capita income and the number of children. Thus, conditional on being in the same vulnerability category and in the same municipality, families with a lower per capita income are included first. Conditional on also having the same per capita income, families with more school-aged children are prioritized. When a family is formally included in the program, the Caixa sets up an account for the family, starts the monthly payments, and issues a magnetic stripe card that allows the family to access the benefits. Importantly, once a family is part of the program, it continues to receive the benefits even if a more deserving family registers in the Cadastro Único. Even if the family's income increases above the threshold for Bolsa Família, the family stays in the program for two more years.¹¹

C. *Bolsa Família's Safeguards against Corruption*

Bolsa Família was designed to minimize the influence of corruption—both at the federal and the local level—and several government agencies operate whistleblowing systems and have the power to investigate alleged abuse of the program.¹²

At the federal level, the main concern is the targeting of funds to municipalities and demographic groups for electoral gains. The risk of geographical quotas being set for electoral purposes is mitigated by tying the process to objective census data. Moreover, the federal gov-

9. The exact categories have changed over time.

10. The process is not standardized, a fact that has repeatedly been criticized by the Federal Court of Accounts (e.g., TCU, 2006). Fortunately, while this introduces an undesirable human element in the allocation of benefits *between* categories, the allocation of benefits *within* categories is unaffected. The estimation strategy outlined in Section III uses only variation within categories.

11. This rule mitigates families' incentives to underreport their income when they update their data. Moreover, due to the frequent increases of the income threshold, many families would otherwise leave the program just to qualify again after the next increase.

12. For a detailed description of Bolsa Família's anti-corruption efforts see Lindert et al. (2007).

ernment has at various points been barred from issuing reports or publications about Bolsa Família in the months leading up to an election.

At the local level, two concerns dominate the efforts: that local officials embezzle funds from the program for personal gain and that they control the allocation process to trade benefits for bribes or votes. The risk of embezzlement is greatly mitigated by transferring funds directly to beneficiaries instead of making bulk payments to state or municipal governments. To address the second concern, the centralized beneficiary selection process was put in place. Moreover, municipalities are required to publish monthly lists of beneficiaries and the payments they received, and all transfers, including full names of the beneficiaries, are also published by the federal government on the Portal da Transparência.

The MDS employs a combination of control mechanisms and incentives for program administrators. Self-reported incomes are compared to records of the department of labor and lists of beneficiaries are cross-referenced with other administrative data such as motor-vehicle registrations. As the majority of beneficiaries is not formally employed, per capita income often cannot be readily verified and income underreporting will often only be detected if a family's lifestyle is incompatible with its reported income—for example, if the family buys a new car. In addition to its verification efforts, the MDS assesses the quality of monitoring and registrations every month and ties federal contributions to administrative costs to an index of municipal management quality, the Índice de Gestão Descentralizada do Município (IGD-M).

In addition to the MDS's internal monitoring initiatives, three other government agencies—the Office of the Comptroller General (CGU), the Federal Court of Accounts (TCU), and the independent public prosecutor's office—are responsible for oversight of the program. The CGU regularly selects municipalities at random to conduct in-depth audits of municipalities.¹³ The CGU's manuals specify the specific actions to be taken concerning Bolsa Família in randomly audited municipalities: First, the auditors will verify the eligibility of a random sample of beneficiaries in the municipality. Second, auditors will cross-reference public employment records with lists of Bolsa Família beneficiaries. Third, payments and operations of the Caixa are scrutinized. Amongst other things, auditors look for proof that benefit cards were delivered to beneficiary families.¹⁴ Fourth, auditors examine monitoring systems for school attendance and compliance with the medical conditionalities. For example, auditors cross-reference reported attendance rates in the online system with entries in class books. Finally, auditors control the existence and recent activities of local governance bodies such as the municipality's social control council. Moreover, the MDS, the CGU, and the Caixa all operate whistleblower systems to report discrepancies and abuse of the program, and auditors investigate specific complaints received through these systems.

13. This is described in more detail in Section IV.

14. This has become less important, as most of the payment cards are now sent directly to beneficiaries.

D. Bolsa Família's Vulnerabilities to Corruption

Despite these efforts, the number of families receiving Bolsa Família exceeds the number of eligible families estimated from the census in a significant share of municipalities (Fried, 2012). While Bolsa Família's centralized beneficiary selection and payment systems are highly effective at preventing outright embezzlement of money designated for beneficiary families, they do not guarantee that the program is effectively administrated on site. Indeed, as others have pointed out (e.g., Lindert et al., 2007), fraud, error, and political interference in Bolsa Família are most likely at the municipal level. As a result, local corruption can affect the effectiveness of the Bolsa Família program by influencing how diligently municipalities fulfill their responsibilities in implementing the program.

A quick Google search reveals numerous news reports of irregularities uncovered by the CGU audits. Common findings include families underreporting their income, schools maintaining sloppy attendance records, social control councils that haven't met for more than a year, delays in the delivery of benefit cards, and the occasional public servant being listed as a member of an unrelated beneficiary household. Note that this implies that local officials can still benefit financially from Bolsa Família, albeit to a much smaller degree than if they could get their hands on all of the payments. Section VI takes a detailed look at the remaining strategies for personal enrichment and shows that the financial damages from these practices are relatively small and cannot account for the differences in program performance.

Bolsa Família's effectiveness in promoting school enrollment depends crucially on its ability to target the families that have the worst expected outcomes in the absence of the cash transfer. In places where corruption is rampant, the program's targeting accuracy is likely to be lower: families, particularly those with an income just above the eligibility threshold, have a strong incentive to underreport their earnings, and corrupt local administrators, aware of this fact, can turn a blind eye in return for a favor. Families whose gross income can be more readily verified will occasionally misrepresent their household composition and include the children of relatives or neighbors to achieve a lower per capita income. In a newsworthy example, a family received monthly benefits for four-year-old Billy da Rosa Silva until a social worker sent to invite the family to a routine health checkup discovered that Billy was a cat (Hider, 2014). There is also evidence that some families deliberately reduce their labor supply leading up to their registration (Firpo et al., 2014).¹⁵

15. Not all mistargeting is the result of intentional deception. Potential beneficiaries might provide weekly instead of monthly wages and mistakes are made because an administrator's fingers miss the mark on a keyboard or sloppy handwriting on a form is incorrectly deciphered when digitizing the data. However, Section VII shows that inadequate administrative processes alone cannot explain the differences in program performance.

III. ESTIMATING THE EFFECTIVENESS OF BOLSA FAMÍLIA

To measure the effectiveness of Bolsa Família in different municipalities, I reconstruct the program's priority strata to find families that were randomly admitted or not admitted, using Bolsa Família's official database and beneficiary selection algorithm.

A. Data—the Cadastro Único

The primary data used for the analysis is the Cadastro Único, the official database used to select beneficiary families for the Bolsa Família program. As the selection algorithm uses only the Cadastro Único, there can be no unobservable characteristics that affect who receives Bolsa Família.

Families with a per capita income of less than half a minimum wage can register in the Cadastro Único and are then required to update their data at least every other year. As registration is voluntary, one might worry that self-selection into the Cadastro Único leads to non-representative estimates. However, this is hardly a concern in practice: for the period of the analysis (2012-2017), the registration rate is close to 100%. So close in fact, that the MDS stopped using the registration rate as a performance indicator in the IGD-M in July 2015 (see Figure A3.6 in Appendix A3.V).

The Cadastro Único provides two types of files on each family. The personal files contain information on each family member's demographics, as well as information on literacy, education, and employment. The family file contains information on the family income, participation in other welfare programs, as well as information on the family's living conditions, such as the material of floors and walls, and access to public services such as water, electricity, sewage systems, and garbage collection. As this additional data is not used to select beneficiaries, I can use it to test whether the identification strategy successfully deals with spurious correlations.

While some data in the Cadastro Único is self-reported by families, other variables are set by the MDS's computer system. For example, the indicator of whether a family benefits from Bolsa Família is automatically changed at some point after the family registered. Thus, a cross-section from the Cadastro Único will contain data from different time-points. I address this problem by considering a family's Bolsa Família status at the beginning of the year. This approach is conservative in that it overestimates the outcomes of supposedly unincluded families, some of which might have benefited from Bolsa Família later in the year.

In addition to the Cadastro Único, this paper uses data from several other sources to investigate how corruption affects the effectiveness of the Bolsa Família program. Data on corruption and Brazil's random audit program is from the Office of the Comptroller General (CGU) and is discussed in more detail in Section IV. The Portal da Transparência publishes expenditures by the Brazilian government, including lists of all the Bolsa Família payments made in any given month and can be used to determine whether some payments have been withheld

because of a failure to comply with the program's conditionalities and whether a family has been excluded from the Bolsa Família program. To better understand the mechanism, the paper uses data from the annual census of social services (the Censo SUAS), quality control data from the index of municipal management quality (IGD-M), and data on denunciations and complaints received through the whistleblowing systems of the MDS.

B. Sample—Randomly Admitted Families

Having access to all the data that the selection algorithm uses, I can reconstruct the priority strata to find families that were randomly included or not included in Bolsa Família. This allows me to estimate the effects of the Bolsa Família program in different years and municipalities.

As Bolsa Família explicitly targets the most impoverished families, there are significant differences between beneficiaries and non-beneficiaries. While some of these differences are observable in the Cadastro Único, others are unobservable and cannot easily be controlled for a regression. Fortunately, because the selection of beneficiaries is based solely on families' information in the Cadastro Único, it cannot be affected by unobservable variables.¹⁶ Thus, having access to both the selection algorithm and the database used to select the beneficiaries, I can identify otherwise identical beneficiaries and non-beneficiaries who have not only the same *observable* characteristics but also the same expected *unobservable* characteristics.

To understand how the selection algorithm can be used to find families that were randomly admitted or not admitted to the Bolsa Família program, recall the four phases of the benefit allocation process: In the *registration phase*, families register in their municipality, and their data is entered in the Cadastro Único. In the *qualification phase*, the Caixa categorizes all eligible families by vulnerability (e.g., indigenous families, families with suspected child labor) and sends the aggregated numbers for each municipality to SENARC. In the *selection phase*, SENARC decides how many benefits it allocates to each vulnerability category and each municipality. In the final *concession phase*, the Caixa determines the actual beneficiaries in each category and municipality based on families' per capita income and the number of children.

The identification strategy closely mirrors the four phases of the beneficiary selection algorithm: First, in line with the *registration phase*, only families that were never part of the program are considered; these are mostly newly registered families and families that did not previously qualify for the program but that now qualify after a change in the eligibility rules of Bolsa Família. Second, mimicking the *qualification phase* and the *selection phase*, families are matched based on the municipality and vulnerability category, and fixed effects are used to exploit only variation between families of the same group. Finally, to capture the *concession phase*, families are additionally matched based on their exact income and their household compo-

16. Lindert et al. (2007, p.45) write: "*Application of eligibility criteria to family data is carried out automatically by the Cadastro Único software, which compares self-reported incomes to the official eligibility thresholds, prioritizing families and assigning benefits according to income and family composition.*"

sition.¹⁷ At each point in time, only the marginal priority strata—those with both selected and unselected families—are considered, leaving only strata where the algorithm randomly included some but not all families in the Bolsa Família program.

While the selection algorithm does not per se prioritize families that have registered earlier, these families have had more opportunities to be included than more recently registered families. The most conservative approach is to include only families that register for the first time, excluding those that only updated data from an earlier registration. This, however, takes quite a toll on the number of families that can be sufficiently precisely matched—especially in smaller municipalities. Alternatively, families can also be matched on the exact month of the registration or the update, but this tends to reduce geographic coverage by a comparable amount. As a compromise, results are shown for all three samples: the most representative sample of all families that can be sufficiently precisely matched irrespective of whether their current information is from a new or an updated registration, the sample of families that are newly registered, and the sample of families that is matched on the exact month of the registration or data actualization.

C. Estimating the Treatment Effects

Assume for a moment that there are only families belonging to the same vulnerability category, with the exact same family income and number of children—i.e., all families are in the same priority stratum—and living in the same municipality. Families are observed for two periods.¹⁸ At $t = 0$, families register (or update their data) and none of the families receive Bolsa Família. Over the next year, the algorithm includes some of the families in the Bolsa Família program, while it cannot accommodate others due to the number of available funds for this category and municipality. Thus, when outcomes are observed again at $t = 1$, some families have been benefiting from Bolsa Família, while others have not.

The educational outcome $Y_{i,f,\theta,m,t}$ of child i in family f of priority stratum θ living in municipality m at time t can be modeled using the following potential outcomes framework:

$$Y_{i,f,\theta,t} = \alpha + \beta \text{Bolsa Família}_{f,t} + X'_{i,f}\gamma_1 + W'_m\gamma_2 + Z'_t\gamma_3 + u_{i,f,\theta,m,t} \quad (3.1)$$

where $\text{Bolsa Família}_{f,t}$ is an indicator of whether family f is included in the Bolsa Família program at time t , $X_{i,f}$ is a vector of (unobservable) child and family characteristics that are fixed over time, W_m is a vector of municipality characteristics, and Z_t are time-varying external factors such as changes in the educational system.

So far I have focused on only one priority stratum. I can estimate the causal effect of Bolsa Família across several marginal priority strata using time \times priority strata fixed effects

17. Note that this is the most conservative approach to address the preferential inclusion of poorer and larger families.

18. Otherwise, repeated re-matching and the possibility of families leaving the program introduce unnecessary complications. However, as shown later, the results are robust if the families are followed for an additional year.

to account for the fact that different strata—and therefore families with different observable characteristics—are on the margin at different points in time. Also, as a child’s family is fixed a specification with child fixed effects can be used to account for observable and unobservable child and family characteristics. This leads to the following specification:

$$Y_{i,f,\theta,m,t} = \beta \text{ Bolsa Família}_{f,t} + \alpha_i + \nu_m + \mu_{\theta,t} + \varepsilon_{i,f,\theta,m,t} \quad (3.2)$$

The treatment variable $\text{Bolsa Família}_{f,t}$ is an indicator that takes value 1 if a family gets included in the Bolsa Família program. The specification includes municipality fixed-effects ν_m and $\text{Year} \times \text{Priority strata}$ fixed effects $\mu_{\theta,t}$ to account for both the randomization within strata and, importantly, also for the fact that different priority strata are on the margin at different points in time. In fact, the priority strata fixed effects non-parametrically control for thousands of combinations of the month of registration, the household’s exact per capita income, and the number of children. The error term $\varepsilon_{i,f,\theta,m,t}$ is allowed to cluster at the family and municipality level to account for the selection of *families* into Bolsa Família and to facilitate comparison with the results in Section IV, where the effect of changes in *municipality*-level corruption are studied. As Bolsa Família is randomly assigned for these families, the coefficient β estimates the causal change in children’s educational outcomes when their families are included in the Bolsa Família program.

D. Validating the Identification Strategy

The identification strategy relies on the fact that the selection algorithm uses only data that is observable in the Cadastro Único so that there can be no unobservable family characteristics that affect which families get included in the Bolsa Família program. Moreover, as the selection algorithm uses only income, household composition, and the vulnerability category, we can think of other data in the Cadastro Único as observable to the researcher, but no to the selection algorithm. This provides a test for whether the identification strategy successfully recovers the random allocation of benefits within priority strata.

Table 3.1 shows that the identification strategy successfully deals with confounding family characteristics that do not influence the selection of beneficiary families but are strongly correlated with per capita income, such as access to utilities, or the materials of the dwelling.¹⁹ Columns (1) and (2) show that these characteristics are relatively well-balanced between future beneficiary and non-beneficiary households in the marginal priority strata. As a more stringent test of balancedness (Pei et al., 2019), Column (3) reports the coefficient when the variable of interest is regressed on an indicator of whether a family will be included in Bolsa Família and the fixed effects, and tests whether these coefficients are jointly significant from zero. A significant

19. The results show the sample of families who have registered for the first time, to ensure that the Bolsa Família indicator is set after the last observed data update.

TABLE 3.1
BALANCEDNESS OF FAMILY CHARACTERISTICS IN MARGINAL PRIORITY STRATA

	No Bolsa Família		Bolsa Família		LHS-test
	(1)		(2)		(3)
	Mean	Standard Deviation	Mean	Standard Deviation	Coeff.
Location: urban	0.810	0.393	0.848	0.359	-0.001
Material: brick	0.754	0.430	0.789	0.408	0.003
Material: clay	0.025	0.156	0.018	0.133	-0.002
Material: timber	0.083	0.275	0.080	0.271	0.002
Material: surplus timber	0.028	0.165	0.030	0.169	-0.001
Material: straw	0.002	0.043	0.002	0.040	0.000
Sewage: canalization	0.394	0.489	0.470	0.499	-0.007 ⁺
Sewage: tank	0.153	0.360	0.144	0.351	0.000
Sewage: tank (rudimentary)	0.264	0.441	0.229	0.420	0.006 ⁺
Sewage: open ditch	0.024	0.152	0.024	0.153	-0.001
Sewage: river	0.008	0.088	0.011	0.103	-0.000
Piped water	0.788	0.408	0.832	0.374	0.007 ⁺
Water source: network	0.671	0.470	0.714	0.452	-0.005
Water source: spring	0.172	0.378	0.159	0.366	0.006
Water source: cistern	0.022	0.148	0.019	0.138	0.001
Indoor bathroom	0.847	0.360	0.883	0.322	0.000
Garbage: collected	0.767	0.423	0.817	0.387	-0.003
Garbage: burned or buried	0.125	0.330	0.099	0.299	0.004
Garbage: dumped on land	0.010	0.097	0.009	0.093	0.000
Garbage: dumped in river	0.001	0.025	0.001	0.023	-0.000
Light: electric (with meter)	0.804	0.397	0.821	0.383	0.001
Light: electric (without meter)	0.060	0.237	0.066	0.248	0.003
Light: oil or gas	0.014	0.118	0.011	0.106	-0.001
Light: candles	0.008	0.088	0.006	0.079	-0.001
F-test: $\chi^2(24)$					23.696
F-test: P-value					0.479
Observations	234757		238733		473490

Notes. This table reports on the balancedness of family characteristics that are not relevant for inclusion in the Bolsa Família program. Columns (1) and (2) present the summary statistics for non-beneficiary and beneficiary families. Column (3) uses a left-hand-side test (Pei et al., 2019) to check whether these characteristics are predictive of a family's inclusion in Bolsa Família in regressions of the form $X_{f,\theta,m,t} = \alpha + \beta BF_{f,t} + \mu_{t,\theta,m} + \varepsilon_{f,\theta,m,t}$. The F-test tests whether these coefficients are jointly different from zero. Significance levels: ⁺ $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

test statistic suggests that these variables are jointly predictive of which families will be included in the program. This is not the case ($\chi^2(24) = 23.696$, $P = 0.479$).

E. Bolsa Família Increases School Enrollment

Before progressing to the main question—whether corruption affects the effectiveness of Bolsa Família—I show that the identification strategy can recover the positive effect of Bolsa Família on school enrollment. Consider the expected impact of a program that pays families for regular school attendance. While we would expect a positive effect, a naïve OLS approach finds a negative effect, because the program was explicitly designed to target families with a low baseline school enrollment. As a final validation of the identification strategy, I thus reestimate the effect of Bolsa Família on school enrollment.

TABLE 3.2
BOLSA FAMÍLIA INCREASES SCHOOL ENROLLMENT

	(1) School enrollment (%) (All marginal families)	(2) School enrollment(%) (Newly registered families)	(3) School enrollment (%) (Same registration month)
BF	1.006*** (0.053)	1.498*** (0.122)	0.919*** (0.067)
Control mean	87.283	86.442	86.809
Child FE	Yes	Yes	Yes
Year \times strata FE	Yes	Yes	Yes
R2	0.925	0.932	0.931
N(municipalities)	5,401	5,068	4,858
N(priority strata)	12,559	8,641	6,008
N(children)	2,573,117	590,630	747,786
N	5,146,234	1,181,260	1,495,572
Years	2012-2017	2012-2017	2012-2017

Notes. This table reports the effect of inclusion into the Bolsa Familia program on children's school enrollment. The dependent variable in all models takes on value 100 if a child is enrolled in school and 0 otherwise. "BF" indicates if a child's family is included in the Bolsa Familia program. Column (1) presents the results for the most representative sample, where families are matched on the municipality, vulnerability category, the exact income, the number of children, and the year the families last updated their data. Column (2) uses the same definition, but only matches families who are newly registered. Column (3) requires families to have last updated their data in the same month. All models include individual-level child fixed effects, municipality fixed effects and "Year \times Priority strata" fixed effects. For the fixed effects, families in each priority stratum have the exact same income, the same number of children, belong to the same vulnerability category, and have last updated their data in the same month. Standard errors are clustered at both the family and the municipality level. Significance levels: $^+ P < 0.1$, $^* P < 0.05$, $^{**} P < 0.01$, $^{***} P < 0.001$.

Table 3.2 shows the change in children's school enrollment if a family gets included in the Bolsa Familia program. Column (1) shows the estimates for all families in marginal priority strata, irrespective whether their information is from a new or updated registration. Column (2) shows the estimates if only families that register for the first time are considered, and Column (3) shows the estimates if families are matched on the month of their registration or last data update. In the first sample, 2,573,117 children are in marginal strata that contain both treated and untreated families. These children live in 5,401 municipalities, covering 97.5% of the municipalities that are eligible for the random audits. If only newly registered families are considered, only 590,630 children from 5,068 municipalities are included in the analysis, and the geographic coverage drops to 91.5%. Likewise, when families are required to have registered in the same month, only 747,786 children are included, and geographic coverage reduces to 4,832 municipalities or 87.7% (see Figure A3.7 in Appendix A3.V for maps of the geographic coverage of each sample).²⁰

Inclusion in Bolsa Família increases enrollment by 1.01 percentage points in the largest sample (Column 1), by 1.50 percentage points in the sample that considers only families who

20. Unsurprisingly, the municipalities that are lost because families cannot be matched precisely enough tend to be less populous. In terms of the population living in municipalities that are eligible for the random audits the coverage is still quite high; 99.7%, 98.7%, and 97.8%, respectively, live in one of the municipalities that are covered by the regressions.

registered for the first time (Columns 2), and by 0.92 percentage points in the sample where families are also matched on the month they registered (Column 3). The effects are highly significant ($P = 0.000$ for all samples) and robust to several alternative specifications (see Appendix A3.II). However, there is considerable heterogeneity in the effectiveness of Bolsa Família across different municipalities (see Figure A3.8 in Appendix A3.V).

These estimates are lower than those of previous evaluations of the program. For example, in one of the earliest studies, Cardoso and Souza (2003) estimate that inclusion in Bolsa Escola, Bolsa Família's predecessor, increases school enrollment by 3 percentage points. Glewwe and Kassouf (2012) find an increase of approximately 5 percentage points using data from 1998 to 2005, Schaffland (2012) documents gains of 4 percentage points using data from 2004 to 2006, and De Brauw et al. (2015) find an effect of 8 percentage points for girls but no significant gains for boys using data from 2009.

Three factors are likely to account for this quantitative difference: School enrollment rates increased considerably over the last two decades, leaving less room for substantial gains. Moreover, the identification strategy uses only the marginal priority strata: as poorer families are more likely to be included with certainty, the marginal strata consist of families with somewhat higher incomes and higher baseline school enrollment. Finally, the use of child fixed effects appears to depress the estimated treatment effects further. Indeed, without controls for individual-level heterogeneity, point estimates are larger (see Table A3.17 in Appendix A3.VI).

IV. LOCAL CORRUPTION AFFECTS PROGRAM EFFECTIVENESS

The first main result of the paper is that local corruption decreases the effectiveness of Bolsa Família, even though local officials are bypassed in both the payment or the beneficiary selection process. Using Brazil's program to randomly audit municipalities as an exogenous shock to corruption (Avis et al., 2018), I show that Bolsa Família's effect on school enrollment increases when municipalities become less corrupt.

A. *Random Audits as Exogenous Shocks to Corruption*

Since 2003, Brazil's federal government operates an anti-corruption program that includes audits of randomly selected municipalities. This audit lottery constitutes a uniquely compelling natural experiment: a series of exogenous shocks, explicitly designed to combat corruption, distributed all over Brazil, and spread over more than a decade. As several municipalities have been audited more than once, it's possible to estimate the effects of a previous audit on instances of corruption detected in later rounds of the program.

In response to rampant abuse of federal transfers by municipal officials, the federal government created the position of the Comptroller General, Controladoria Geral da União (CGU) in 2003. In the same year, the Programa de Fiscalização por Sorteios Públicos started to randomly

select municipalities for comprehensive audits in a draw that is held in conjunction with the national lottery. Municipalities with less than 500,000 are eligible for the lottery, whereas more populous cities are subject to other audits.²¹ Once a municipality has been audited, it cannot be selected again for some time.²²

If a municipality is randomly selected, the CGU lists all federal transfers made to this municipality in the previous years and randomly chooses a number of them for in-depth audits. The CGU then issues an inspection order for each of the selected transfers and sends a team of 10-15 auditors to the municipality, usually within less than a month after the lottery. Auditors are highly qualified and competitively paid and, consequently, less susceptible to bribery than other public employees (Avis et al., 2018). The auditors carefully examine expenditures associated with the inspection order, verify the delivery of goods and services paid for by the transfer, and check whether procurement and hiring decisions comply with the relevant laws. Auditors also engage with the local community and with municipal councils to gather additional information and register complaints (Ferraz and Finan, 2008). Once completed, the auditors share their findings with federal prosecutors, the federal police, the local judiciary, and the city council.

Since round 17 of the audit lottery, the CGU focuses on specific sectors in each draw.²³ As not all sectors are audited in every round, one might worry that some audits don't have the same corruption reducing effect on the Bolsa Família program. Fortunately, Bolsa Família was subject to every round of the program since 2011.

Moreover, while the inclusion of all audits may underestimate the true impact on the Bolsa Família program, there are several reasons why one should still expect an effect of the audits: First, although auditors are not allowed to venture into other sectors, they report additional suspicions back to the CGU, which can then take appropriate actions. Second, even though the audited sectors differ, the people implicated will often be the same, as many municipalities are dominated by a small political elite. Third, using evidence from municipalities that have been audited more than once, Avis et al. (2018) show that, if anything, the corruption reduction is somewhat weaker in the audited sector, suggesting that politicians assume that a sector is now less likely to be audited again. Finally, there are spill-over and learning effects between municipalities through local media and shared political networks (Avis et al., 2018) and it is reasonable to expect that these effects are at least as important within a municipality. Indeed, Table A3.18 in Appendix A3.VI shows that corruption in the Bolsa Família program and the educational system more generally are highly correlated with corruption in other sectors.

21. This affects 31 municipalities, mostly state capitals, home to approximately 27% of the Brazilian population.

22. The exact rules have varied over time. See Ferraz and Finan (2008) and Avis et al. (2018) for details of the audit lottery program.

23. In round 27, for example, the auditors looked at transfers related to social assistance, agriculture, commerce and services, and culture in municipalities with a population of more than 100,000, and at all these sectors, plus health and education-related transfers in municipalities with a population between 20,000 and 100,000 thousand. In municipalities with less than 20,000 inhabitants, all sectors were audited.

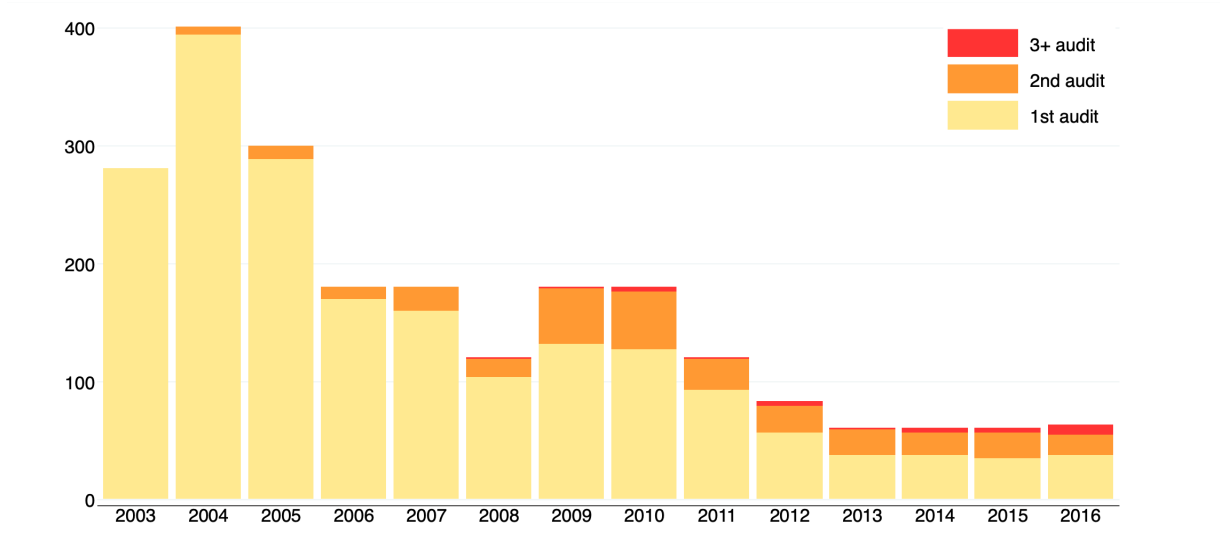


FIGURE 3.1
Timeline of Random Audits

Notes. This figure displays the timeline of random audits under 40 rounds of the Programa de Fiscalização em Entes Federativos (2003-2015) and the random third cycle of its successor, the Programa de Fiscalização em Entes Federativos (2016). Colors indicate whether a municipality is being randomly audited for the first time, the second time, or at least the third time.

In 2016, the CGU was reconstituted as the Ministério da Transparência, Fiscalização e Controladoria-Geral da União and the random audit lotteries were superseded by the Programa de Fiscalização em Entes Federativos. While some audits remain random, the new program also conducts non-random audits. So far, only the third cycle was lottery-based,²⁴ while the first and fourth cycle used insights from the previous program to select those municipalities deemed to be the most vulnerable and the second cycle conducted comprehensive audits of state capitals.

B. Validating the Corruption Reduction after Random Audits

As several municipalities have been audited more than once, it's possible to estimate the effects of a previous audit on instances of corruption and mismanagement detected in later rounds of the program (Avis et al., 2018). Since 2003, 1,956 municipalities with a population of less than 500,000 were randomly audited; 285 of which were audited twice, 25 three times and one municipality four times—2,267 random audits in total (Figure 3.1).

The treatment variable is an indicator, *Past audit*, whether a municipality has been randomly audited in the past. For the construction of the treatment variable, I consider all municipalities with a population of less than 500,000 in the census of the year 2000 and I include all 40 rounds of the original Programa de Fiscalização por Sorteios Públicos as well as the random

24. Eligibility rules differed somewhat from previous lotteries. As a safeguard, I include only municipalities that would have been eligible under the previous regime. Appendix A3.II shows that the results are robust if only the original program is considered.

third cycle of its successor, the Programa de Fiscalização em Entes Federativos.²⁵ As a result, the sample differs from the one used by Avis et al., who restrict their analysis to rounds 22 to 38 of the audit lottery program to focus on the two electoral terms from 2004 to 2012. Moreover, in round 36, the audits of several selected municipalities were canceled less than three weeks later due to a strike of the auditors.²⁶ Avis et al. code these municipalities as having been treated.²⁷

To validate the corruption-reducing effect of previous audits in the full sample, I re-estimate Equation (16) in Avis et al. (2018):

$$\begin{aligned} \log(\text{Corruption}_{m,s,t}) = & \alpha + \beta \text{ Past audit}_{m,s,t} + Z'_{m,s,2000}\gamma \\ & + f(\text{Inspection orders}_{m,s,t}) + \nu_s + \mu_t + \varepsilon_{m,s,t} \end{aligned} \quad (3.3)$$

where the logarithm of the number of detected incidents in municipality m in round t of the audits program, is regressed on the treatment variable $\text{Past audit}_{m,s,t}$, that takes value 1 if a municipality has been subject to a random audit in an earlier round of the program. Depending on the specification, the model controls for socioeconomic factors $Z_{m,s,2000}$: the logarithm of population, the share of population living in an urban area, income inequality, log. income per capita, and illiteracy rate—measured in 2000 before the inception of the audit program. Because the number of transfers and programs auditors inspect, *Inspection orders*, directly affects the number of uncovered incidents, it is controlled for either logarithmically or non-parametrically. State fixed effects ν_s mirror the stratification of the lottery: different locations face different probabilities of being audited depending on the number of municipalities in a state. Finally, fixed effects for the round of the audits program μ_t account for the fact that the number of municipalities that have been *previously* audited is necessarily weakly increasing over time and that the sectors chosen for the audits vary across rounds. With the appropriate fixed effects in place, the coefficient β can be interpreted as the causal effect of a previous audit on corruption.

Table 3.3 displays the effects of having previously been audited at random on the total number of irregularities, cases of mismanagement, and corruption²⁸ in three different specifications: using the logarithm of the number of inspection orders without controlling for sociodemographic factors (Columns 1, 4, and 7), adding the sociodemographic factors (Columns 2, 5, and 8), and including fixed effects for the number of inspection orders and control variables—the preferred specification of Avis et al. (Columns 3, 6, and 9).

25. See Figures A3.9 and A3.10 in Appendix A3.V for the geographic distribution of the audits and the *Past audit* variable over time.

26. *Portaria n° 1.713* (August 10, 2012)

27. Appendix A3.II shows that the results are robust to alternative definitions of the treatment variable.

28. Since round 20 of the program, the CGU has coded the severity of its findings internally as either *falha formal*, *falha média*, or *falha grave*—formal, moderate, and severe cases. As discussed by others (Avis et al., 2018; Zamboni and Litschig, 2018), the distinction between moderate and severe cases is primarily a question of the potential financial damage and says relatively little about the nature of the corrupt action. There is also considerable overlap between losses judged as moderate and severe (see Figure A3.11 in Appendix A3.V); thus I use the classification of Avis et al. (2018) and refer to formal errors as instances of *mismanagement* and to moderate and severe findings as acts of *corruption*.

TABLE 3.3
RANDOM AUDITS SIGNIFICANTLY DECREASE CORRUPTION

	Irregularities (Log.)			Mismanagement (Log.)			Corruption (Log.)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Past audit	-0.044* (0.020)	-0.049* (0.020)	-0.047* (0.020)	-0.070 (0.051)	-0.067 (0.051)	-0.074 (0.051)	-0.040+ (0.022)	-0.046* (0.022)	-0.043* (0.022)
Population (Log.)		0.028** (0.010)	0.033** (0.010)		-0.062* (0.026)	-0.070** (0.027)		0.040*** (0.011)	0.046*** (0.011)
Income inequality (Gini)		0.158 (0.122)	0.215+ (0.123)		-0.036 (0.363)	-0.186 (0.371)		0.188 (0.136)	0.277* (0.139)
Income per capita (Log.)		-0.090* (0.036)	-0.120*** (0.036)		0.083 (0.096)	0.104 (0.095)		-0.101** (0.039)	-0.133*** (0.039)
Illiteracy		0.002+ (0.001)	0.002 (0.001)		0.002 (0.004)	0.002 (0.004)		0.003* (0.002)	0.003+ (0.002)
Urban population		0.101* (0.050)	0.112* (0.051)		0.180 (0.129)	0.177 (0.132)		0.097+ (0.056)	0.111* (0.057)
Constant	0.868*** (0.106)	0.949*** (0.223)	3.840*** (0.223)	-1.002*** (0.219)	-1.068+ (0.585)	1.258* (0.627)	0.808*** (0.121)	0.813*** (0.246)	3.657*** (0.250)
Inspection orders	Log. Yes	Log. Yes	Nonpar. Yes	Log. Yes	Log. Yes	Nonpar. Yes	Log. Yes	Log. Yes	Nonpar. Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Lottery FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.804	0.807	0.823	0.275	0.278	0.307	0.780	0.785	0.801
F	72.829	64.807	18.741	19.494	17.047	8.673	58.480	52.924	17.473
N	1432	1432	1432	1432	1432	1432	1432	1432	1432

Notes. This table replicates the effect of having been randomly audited in the past on uncovered instances of corruption and mismanagement in later rounds reported in Avis et al. (2018). The dependent variable in Columns (1) to (3) is the logarithm of the total number of irregularities uncovered by the random audit program. Columns (4) to (6) include only instances of mismanagement (*falha formal*), and Columns (7) to (9) only instances of corruption (*falha média* or *falha grave*). “Past audit” indicates that a municipality has been audited at random in a previous round of the program. “Population (Log.)”, “Income inequality (Gini)”, “Income per capita (Log.)”, and the rates of “Illiteracy” and “Urban population” control for municipality characteristics in 2000, before the inception of the audits program. Columns (3), (6), and (9) include fixed effects for the number of inspection orders issues, while the other columns control for the logarithm or inspection orders. All models include fixed effects for the state and the round of the audit lottery. Robust standard errors are reported in brackets. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

As in the restricted sample used in Avis et al. (2018), there is a clear difference between mismanagement and corruption. For mismanagement, there is no effect of previous audits and, other than population, none of the control variables has any predictive power. In contrast, there are significantly fewer instances of corruption in audited municipalities, although the reduction is slightly less pronounced in the full sample.

C. Estimating the Change in Treatment Effects

To estimate the change in the effectiveness of Bolsa Família after a municipality has been randomly audited, Equation (3.2) can be appended with the audit indicator and an interaction term:

$$Y_{i,f,\theta,m,t} = \beta \text{ Bolsa Família}_{f,t} + \gamma \text{ Past audit}_{m,t} + \delta (\text{BF} \times \text{Past audit})_{f,m,t} + \alpha_i + \nu_m + \mu_{\theta,t} + \varepsilon_{i,f,\theta,m,t} \quad (3.4)$$

Both Bolsa Família and the audits are randomly assigned (once we correctly account for stratification), and we can interpret the coefficient δ as the causal change in the effectiveness of Bolsa Família after a municipality has been audited at random. All standard errors are clustered at both the family and the municipality level, to account for the selection of *families* into Bolsa Família and the selection of *municipalities* in the audit lotteries.

D. Bolsa Família Is More Effective after a Random Audit

After a random audit, Bolsa Família's effect on school enrollment increases. This finding is robust to a large number of alternative specifications.

Figure 3.2 shows that the school enrollment gains from inclusion in the Bolsa Família program increase significantly after a municipality has been audited at random. In the most representative sample, Bolsa Família is estimated to increase school enrollment by 0.90 percentage points in unaudited municipalities, but by 1.18 percentage points in audited municipalities ($P = 0.031$ for the interaction term). Thus, a random audit increases the effect of Bolsa Família by 31%. The estimates are somewhat larger if only families who registered for the first time are considered. Bolsa Família is estimated to increase school enrollment by 1.31 percentage points in unaudited municipalities, but by 1.82 percentage points in audited municipalities, an increase of 39% ($P = 0.032$ for the interaction term). In the sample that matches families on the month of registration, Bolsa Família is estimated to increase school enrollment by 0.81 percentage points before an audit. The effect increases to 1.09 percentage points after a municipality has been audited at random ($P = 0.070$ for the interaction term), a gain of about 34% relative to the pre-audit level.

The effect is robust for several alternative specifications. For comparison, Column (1) of Table 3.4 shows the initial estimate.²⁹ The standard estimation follows each child for just two

29. The table shows the estimates for the most representative sample. However, the result is also robust

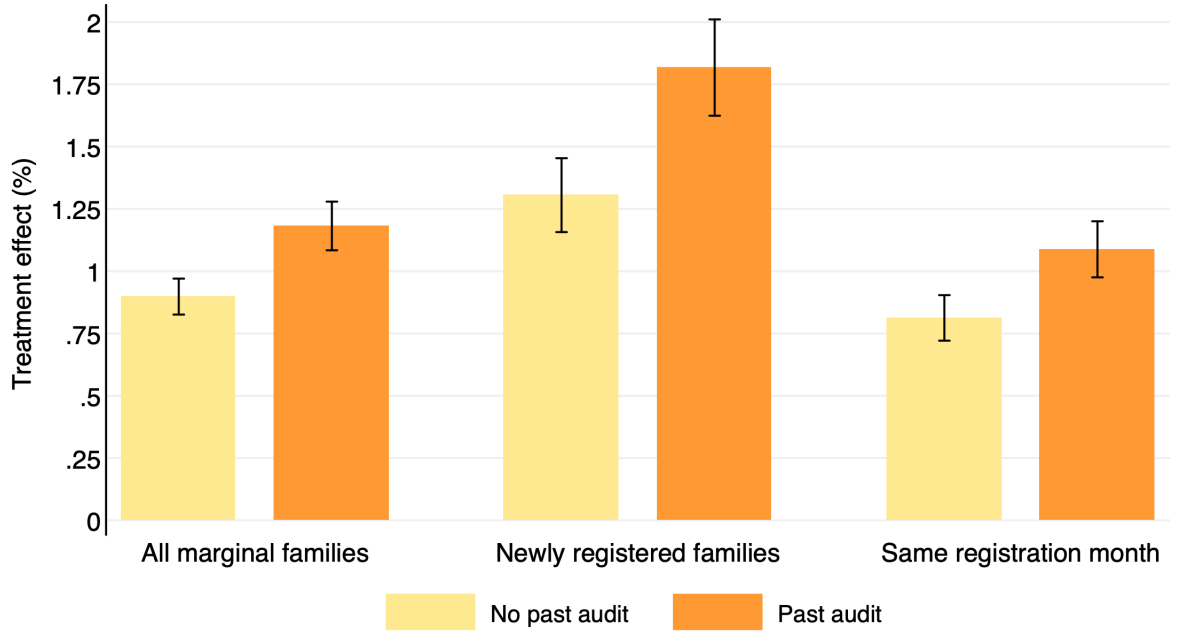


FIGURE 3.2

Bolsa Família Is More Effective after a Random Audit

Notes. This figure displays the effect of inclusion into the Bolsa Família program and random audits on children's school enrollment. Treatment effects are estimates using the specification in Equation (3.4). Colors indicate whether a municipality has been audited at random.

years to mitigate the problem of included families dropping out of Bolsa Família or unincluded families gaining access to the program. However, it is conceivable that the beneficial effect of the random audits is less pronounced after families have been included for some time. Column (2) displays the estimates of an intent-to-treat approach that follows families for an additional year and ignores (potentially non-random) dropout and new inclusions. As expected, ignoring later entries and exists to the program leads to lower estimates of Bolsa Família's effectiveness; 0.44 percentage points as opposed to 0.90 percentage points. However, the gains after a random audit are slightly larger and continue to be statistically significant ($P = 0.045$).

Educational participation varies by age and gender, and there are several well-documented patterns such as the delayed enrollment of younger children and the increased dropout rate for older boys (De Brauw et al., 2015). Column (3) shows that the result is robust if Age \times Sex fixed effects are included. The interaction term continues to be significant ($P = 0.044$), and the estimates of Bolsa Família's impact are similar to the initial result—both before and after a municipality is audited at random.

Bolsa Família has a special provision for children above the age of 15. These children have a lower attendance requirement (75% instead of the usual 85%) and are legally able to work as

for all specifications in the sample of newly registered families and for almost all specifications in the sample of families that are also matched on the month they updated their data. See Appendix A3.II for details on the robustness checks, including the results in the other samples, and additional tests.

TABLE 3.4
ROBUSTNESS TO ALTERNATIVE SPECIFICATIONS

	School enrollment (%)									
	(1) Original specification	(2) Including third year	(3) Demographic fixed effects	(4) Including teenagers	(5) Inverse prob. weights	(6) Propensity score 10–90%	(7) Only original audit lottery	(8) Including current audit	(9) Data from current year	(10) Excluding 2017
BF	0.898*** (0.072)	0.442*** (0.084)	0.992*** (0.071)	0.750*** (0.063)	0.922*** (0.067)	0.876*** (0.072)	0.897*** (0.072)	0.900*** (0.073)	0.509*** (0.062)	0.915*** (0.076)
Past audit	-0.157 (0.281)	-0.588* (0.245)	-0.114 (0.265)	-0.095 (0.255)	-0.104 (0.292)	-0.096 (0.291)	-0.043 (0.352)	-0.648** (0.240)	-0.119 (0.293)	-0.044 (0.352)
BF × Past audit	0.283* (0.131)	0.366* (0.182)	0.258* (0.128)	0.231* (0.113)	0.274* (0.128)	0.263* (0.132)	0.287* (0.131)	0.275* (0.130)	0.252* (0.119)	0.292* (0.140)
Control mean	87.283	87.920	87.283	87.214	87.283	87.211	87.283	87.277	88.747	87.152
Child FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age × Sex FE	No	No	No	No	No	No	No	No	No	No
Year × strata FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.925	0.858	0.928	0.929	0.931	0.926	0.925	0.925	0.925	0.923
N(municipalities)	5,401	5,401	5,401	5,416	5,401	5,401	5,401	5,401	5,395	5,381
N(priority strata)	12,559	12,752	12,559	15,250	12,559	12,547	12,559	12,559	7,233	11,696
N(children)	2,573,117	2,585,404	2,573,117	3,058,499	2,573,117	2,127,141	2,573,117	2,573,117	2,266,681	2,384,393
N	5,146,234	7,171,463	5,146,234	6,116,998	5,146,234	4,254,282	5,146,234	5,146,234	4,533,362	4,768,786
Years	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2016

Notes. This table reports on the robustness of the effect of inclusion into the Bolsa Família program and random audits on children's school enrollment. The dependent variable in all models takes on value 100 if a child is enrolled in school and 0 otherwise. "BF" indicates if a child's family is included in the Bolsa Família program, "Past audit" indicates that a municipality has been audited at random, and "BF × Past audit" is the interaction of the two treatments. Column (1) presents the results from the initial specification for comparison. Column (2) presents the results if children are followed for up to three years. Column (3) presents the results if "Age × Sex" fixed effects are included to control for different patterns of educational participation. Column (4) presents the results if teenagers up to age 17 are included. Column (5) presents the results if stabilized inverse probability weights are applied to correct for differences in the treatment propensity. Column (6) presents the results if families with a treatment propensity of less than 10% or more than 90% are excluded. Column (7) presents the results if only the 40 rounds of the original Programa de Focalização por Sorteios Públicos are considered. Column (8) presents the results if municipalities are considered to be previously audited if the audit occurs in the same year. Column (9) presents the results if only families are included that have updated their data in the current year. Column (10) presents the results when the year 2017, where information is observed mid-year, is excluded. All results are for the most representative sample, where families are matched on the municipality, vulnerability category, the exact income, the number of children, and the year the families last updated their data. All models include individual-level child fixed effects, municipality fixed effects and "Year × Priority strata" fixed effects. For the fixed effects, families in each priority stratum have the exact same income, the same number of children, belong to the same vulnerability category, and have last updated their data in the same month. Standard errors are clustered at both the family and the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

part of an apprenticeship. Column (4) shows that although the point estimates are somewhat smaller if older children are included, Bolsa Família continues to be significantly more effective after a random audit ($P = 0.040$).

Although families within a priority stratum are randomly admitted to Bolsa Família, families in some strata have considerably higher probabilities of being included (see Figure A3.1 in Appendix A3.II). To address this, Column (5) tests whether the result is robust if stabilized inverse probability weights are applied to correct for the higher treatment propensities in some strata.³⁰ The estimates are of similar magnitude, and the interaction term continues to be significant ($P = 0.032$). Column (6) tests whether the effect persists if families with a treatment propensity of less than 10% or more than 90% are excluded from the analysis. The estimates are again of similar magnitude, and the interaction term continues to be significant ($P = 0.046$).

In 2016, the Programa de Fiscalização por Sorteios Públicos was superseded by the Programa de Fiscalização em Entes Federativos. As the third cycle of the new program consisted of a random audit lottery, it is included in the definition of the *Past audit* indicator. Column (6) shows that the result is robust if only the 40 rounds of the original audit lottery are considered. The estimates are again of similar magnitude, and the interaction term continues to be significant ($P = 0.029$).

Because the audit reports are often only published late in the calendar year or even at the beginning of the next one, the *Past audit* indicator is defined to take value 1 if a municipality has randomly been audited in a *previous* year. Column (7) relaxes this and considers municipalities as having been audited in the past, even if the audit takes place in the current year. Despite reclassifying 198 municipalities, the result does not change significantly: Bolsa Família is again roughly 30% more effective after a municipality has been audited at random and the interaction continues to be significant ($P = 0.035$).

Families are required to update their Cadastro Único registration at least every other year. Thus, in a given year, the data of some families actually reflects information from previous years. So far, this has been addressed by constructing priority strata so that families that potentially have outdated information are in separate priority strata. However, Column (9) shows that the result is robust if these priority strata are excluded ($P = 0.034$).

Finally, the data for this paper were obtained in late 2017, so the last year of the Cadastro Único data represents the state in June 2017. As a result, families that registered towards the end of the sample had less time to realize their gains, although the year fixed effects mitigate this problem to some degree. Column (10) shows that the result persists if data from 2017 is excluded: Bolsa Família is again roughly 30% more effective after a municipality has been audited at random and the interaction continues to be significant ($P = 0.037$).

30. The weights take the form $w_{1,f} = \frac{Prob(BF)}{Prob(BF|\theta,m)}$ for families that get included and $w_{0,f} = \frac{1-Prob(BF)}{1-Prob(BF|\theta,m)}$ for families that don't get included, where $Prob(BF | \theta, m)$ denotes the conditional probability of being included in Bolsa Família for a family in priority stratum θ and municipality m .

V. UNDERSTANDING THE MECHANISM

Because municipalities are responsible for registering potential beneficiaries, local corruption is most likely to play a role at the registration stage. This section shows how local corruption increases the chance that ineligible families benefit from Bolsa Família—using a theoretical model, administrative data, and a field experiment—and how this mistargeting decreases the effectiveness of the program.

Throughout the following sections, statements from the audit of Sete Quedas (MS) in March 2015 are used to illustrate the typical ways in which municipalities fall short of their responsibilities, before testing whether these factors are likely to explain the observed performance gains. With 10,780 inhabitants at the time of the last census, Sete Quedas ranks 2,821 out of 5,570 municipalities—very close to the median. However, the municipality is chosen not so much for its demographic representativeness, but rather because the audit report is a relatively comprehensive summary of findings that are frequently encountered in the audits of other municipalities. It is worth noting that the municipality punches way below its weight in terms of educational achievement; it ranks 5,290 out of 5,570 in school enrollment despite average salaries being in the top quintile of the country.

A. *Income Underreporting and Mistargeting*

"We found beneficiaries of the Bolsa Família program with a per capita income higher than that established in the program's legislation." (CGU, 2015)

Bolsa Família targets those families that are most likely to underinvest in human capital: low-income families, families with many children, and families in marginalized groups. The program's impact depends crucially on its ability to reach these families. High levels of corruption make the program vulnerable to exploitation by families that don't qualify under the rules of the program. As the audit report puts it, "underreporting of income during the registration in the Cadastro Único [...] may lead to undue receipt of benefits by families outside of the target audience of government social programs and non-treatment of families in the target audience" (CGU, 2015).

This suggests a straight forward explanation why Bolsa Família should be more effective after a random audit: families that underreported their income are excluded from the program and, going forward, the municipality pays closer attention to families' incomes during the registration process. As a result, Bolsa Família is more likely to reach the families that benefit the most.

B. A Model of Income Underreporting

To better understand how income underreporting affects Bolsa Família's effectiveness, consider a simplified model of the registration process, where families decide what income to report, and the families with the lowest self-reported income are included in the program. If a family's reported income deviates too much from its true income, it risks being caught. A key assumption of the model is that the risk of detection is lower in high corruption municipalities. As more families underreport their income, Bolsa Família can no longer target the families that benefit most, and its effectiveness decreases.

The model's structure is as follows: First, SENARC decides how many families to include in Bolsa Família, based on income data from the census. After that, families register in the Cadastro Único, potentially underreporting their incomes. The Caixa then includes the families with the lowest incomes until all places are filled. Finally, a family that is included but has underreported its income may be detected and face the consequences.

At the beginning, SENARC decides how many families to include in Bolsa Família. Let there be N families of the same vulnerability category living in the same municipality. Because SENARC sets separate numbers of beneficiaries for each category and municipality, it's reasonable to focus on just one such group when considering the strategic motives. It's assumed that the number of places M allocated by SENARC is such that not all families will be covered, $M < N$. There are two possible interpretations of this assumption: there could be insufficient funds to cover all families or the census implies that fewer than N eligible families live in the municipality.

In the next step, families register in the Cadastro Único and decide what income y they report. Families with an income below the eligibility threshold \bar{y} qualify for Bolsa Família. Each family knows its own true per capita income x and the true distribution of income per capita, modeled by the cumulative distribution function $F(x)$ over the closed interval $[0, \bar{x}]$. Thus, no family has a negative income, but some families may have zero income.

The Caixa then selects the M families that report the lowest incomes below the eligibility threshold \bar{y} from the Cadastro Único and pays them a fixed benefit of $b > 0$.

Once included in Bolsa Família, there is a risk of detection if a family underreported its income, which is lower in high corruption municipalities. Let's assume that underreporting one's income comes at an expected cost $c_m \cdot (x - y)$ that is increasing in the difference between the true income x and the reported income y . This captures several possible mechanisms, for example, that families are more likely to be found out if their true incomes—and their lifestyles—differ more from their reported incomes or that the punishment for underreporting is proportional to the deviation from the true income. The key assumption is that c_m is lower in municipalities where corruption is prevalent, either because the probability of being found out is smaller or

because the expected punishment is less severe for a given degree of underreporting.³¹

Thus, when family i with income x_i registers in municipality m , it decides what income y_i to report to maximize the expected utility:

$$\max_{y_i \in [0, x_i]} U(y_i | x_i) = \text{Prob}(\text{receiving BF} | y_i) (b - c_m \cdot (x_i - y_i)) \quad (3.5)$$

Proposition 1 *The optimal reporting function for a family with income x_i is given by*

$$y(x_i) = \begin{cases} 0, & x_i \leq x^* \\ x_i - \frac{b}{c_m} + \frac{1}{\mathfrak{B}(M-1; N-1, F(x_i))} \int_{x_i}^{x^r} \mathfrak{B}(M-1; N-1, F(\alpha)) d\alpha, & x^* < x_i \leq x^r \\ x_i, & x^r < x_i \end{cases} \quad (3.6)$$

where $\mathfrak{B}(m; n, p)$ denotes the cumulative binomial distribution that a binary event with probability p occurs at most m out of n times and $x^r = \bar{y} + \frac{b}{c_m}$ denotes the highest income that allows families to benefit from underreporting their income. The function exhibits a discontinuity at $x^* \in \left[0, \frac{b}{c_m}\right]$ that is defined by the following equation:

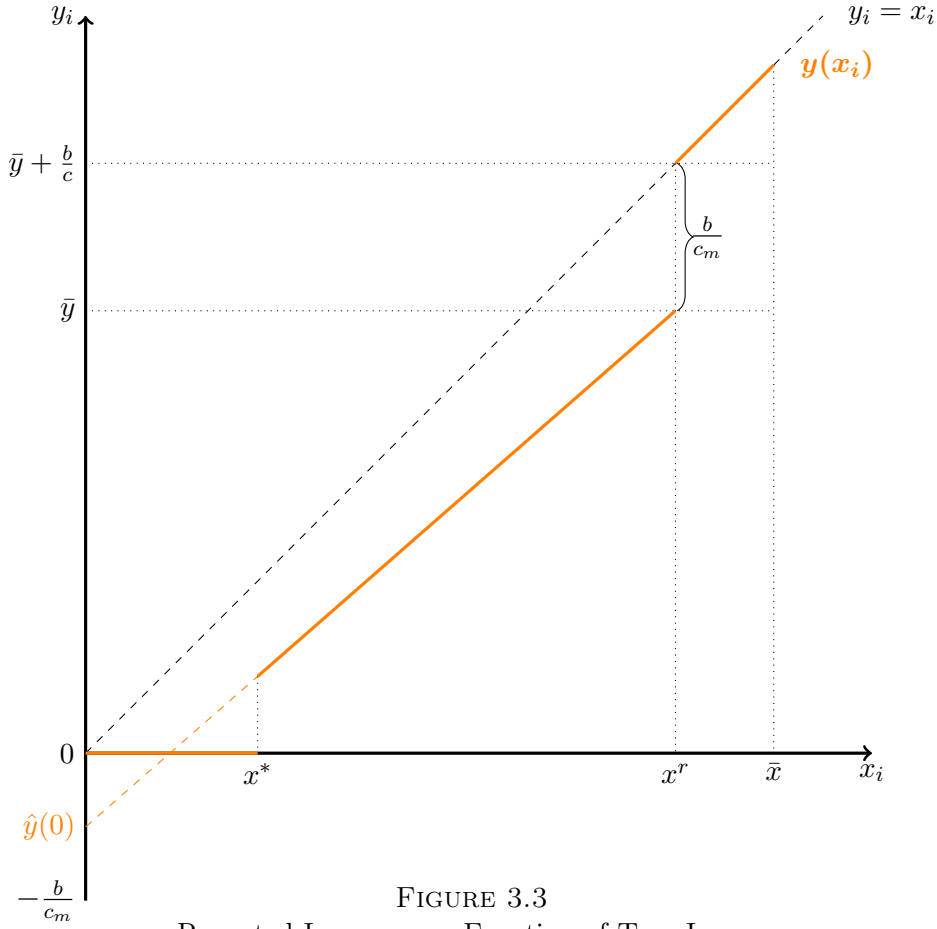
$$\frac{b}{c_m} = x^* + \frac{\int_{x^*}^{x^r} \mathfrak{B}(m-1; N-1, F(\alpha)) d\alpha}{\mathfrak{B}(M-1; N-1, F(x^*)) + \frac{M}{NF(x^*)} (1 - \mathfrak{B}(M; N, F(x^*)))} \quad (3.7)$$

Proof. See Appendix A3.I.³²

Figure 3.3 illustrates the optimal reporting function, which has several intuitive properties. The amount of underreporting is stronger if the potential benefits b are higher, and it is lower if the expected costs c_m of being found out increase when a municipality becomes less corrupt. The model also predicts that families at the lower end of the income distribution and families with an income close to the eligibility threshold have the strongest incentives to underreport their income. Specifically, the model predicts that a disproportionate number of families report an income of zero—in line with the observed distribution of income in the Cadastro Único—and that more (fewer) families report an income of zero or an income that makes them eligible for Bolsa Família if the benefits b (costs c_m) increase (see Appendix A3.I).

31. Because only one municipality is considered in solving the model, the subscript is omitted in the proofs to simplify notation: $c = c_m$.

32. The model resembles an auction where families underbid each other to win one of the places in the Bolsa Família program. As a result, the proof is related to work in the theory of procurement auctions (e.g., Calveras et al., 2004; Compte et al., 2005; Li and Zheng, 2009) and auctions with capped bids (e.g., Che and Gale, 1998; Zheng, 2001; Gavius et al., 2002; Chen and Chiu, 2011). For convenience, it is assumed that families with an income above $x^r = \bar{y} + \frac{b}{c}$ report truthfully. However, none of the model's predictions depend on this assumption. Any reporting behavior such that the family never qualifies for Bolsa Família, i.e., $y_i > \bar{y}$, is an equilibrium. Thus, the second discontinuity at x^r in Figure 3.3 is not necessarily there, as families could in principle report an income arbitrarily close to \bar{y} . See Claim 1 in Appendix A3.I.



Notes. This figure illustrates the relationship between a family's true income x_i and the income it reports y_i . The solid orange line is the optimal reporting function $y(x_i)$ if families are constrained to report non-negative incomes. The dashed orange line is the unconstrained optimal reporting function $\hat{y}(x_i)$. The dashed black line describes truthful reporting, $y_i = x_i$.

The model predicts that the true income of supposedly eligible families is higher than the income reported in the Cadastro Único and that the difference is more pronounced in places with more corruption, where families are less likely to be detected. Thus, even if the poorest M families are included in each municipality, a family with a reported income of y in a place with little underreporting is likely to be poorer than a family with the same reported income in a high corruption municipality where underreporting is more prevalent.

If Bolsa Família has a stronger effect on families with a lower *true* income, the model predicts that, conditional on the *reported* income, the treatment effects are smaller if there is more underreporting. Suppose that school enrollment is increasing with diminishing returns in the true family income³³ including potential Bolsa Família benefits then:

33. This captures the empirical relationship between income and educational participation reasonably well, see Figure A3.13 in Appendix A3.V.

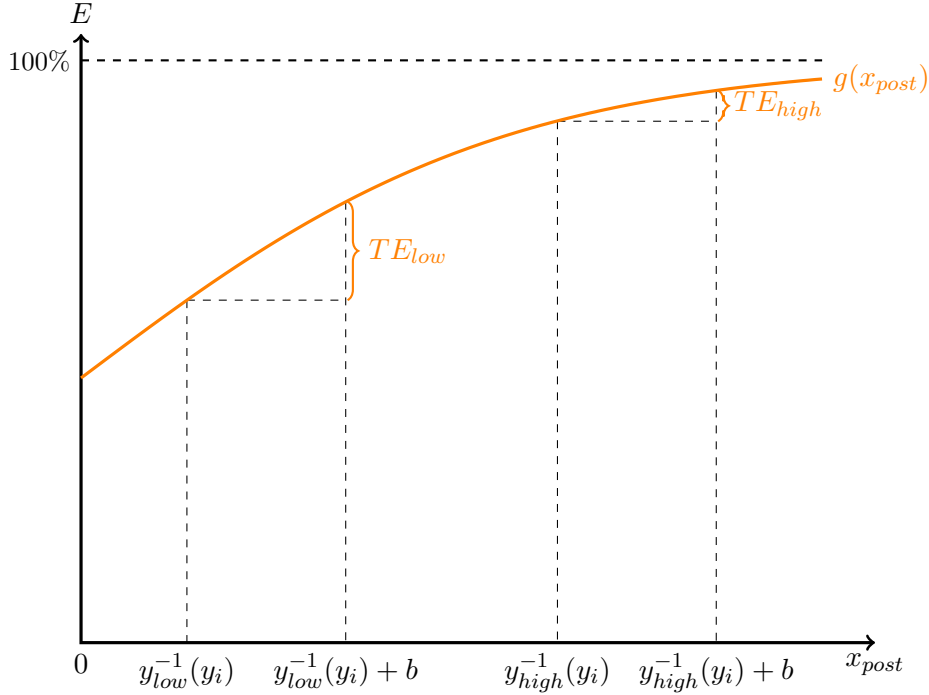


FIGURE 3.4

Predicted Treatment Effects for Low and High Rates of Underreporting

Notes. This figure displays the relationship between the extent of income underreporting and the predicted treatment effects of Bolsa Família. x_{post} denotes the true income after the treatment assignment, i.e., $x_{post,i} = x_i + b$ if family i receives Bolsa Família and $x_{post,i} = x_i$ otherwise. $y_{high}^{-1}(y_i)$ is the true income of a family that reports an income of y_i in a municipality with a high degree of income underreporting. $y_{low}^{-1}(y_i)$ is the true income of a family that reports an income of y_i in a municipality with a low degree of income underreporting. The function $g(x_{post})$ describes the relationship between school enrollment and income.

Proposition 2 Let x_{post} be the true income after the treatment assignment, i.e., $x_{post,i} = x_i + b$ if family i receives Bolsa Família and $x_{post,i} = x_i$ otherwise. Suppose that the function $g(x_{post})$ describes the relationship between school enrollment and income and that it is twice continuously differentiable with $g' > 0$ and $g'' < 0$. For $y_i > y(x^*)$:

1. The baseline school enrollment, $g(y^{-1}(y_i))$, is increasing in c .
2. The treatment effect on school enrollment, $g(y^{-1}(y_i) + b) - g(y^{-1}(y_i))$, is decreasing in c .

Proof. The first claim follows from the fact that $y(x_i)$ is decreasing in c . The second claim additionally uses the concavity of $g(x_{post})$. \square

Figure 3.4 illustrates the effect of underreporting on the expected treatment effects, conditional on the reported income y .

Previously, the assumption was that all families underreport their incomes equally. If some families are less able or willing to underreport, it needs no longer be the case that the M families

that end up benefiting from the program are indeed the poorest M families.³⁴ If this is the case, the true income of some included families will exceed the true income of some families that did not underreport to the same degree. As a result, the expected true income of the included families will be even higher.

There are many situations where this is likely to be the case. For example, it is easier to verify the income of a household with a member who is formally employed or receives additional social assistance payments. Similarly, if a home visit is conducted to register some but not all of the families, social workers might be able to judge the true income of the visited families more accurately. Local corruption can compound the problem if some families have friends or relatives with some influence on the income verification process. Moreover, understanding and gaming the system requires information about its rules and is cognitively demanding, so some families might not be aware of the strategic component.³⁵ Finally, families and communities might have different social norms about claiming government benefits despite not being entitled to them.³⁶

C. Testing the Model in the Administrative Data

The registration model makes an easily testable prediction about the distribution of reported incomes as a municipality becomes less corrupt. In the aftermath of a random audit, fewer families should report an income that qualifies them for Bolsa Família, and fewer families should report an income of zero (see Claims 10 and 12 in Appendix A3.I).

Table 3.5 shows that this is indeed the case. The share of families in the Cadastro Único³⁷ reporting an income below the eligibility threshold at the time of their registration falls by 1.09 percentage points (Column 1) and the share of families claiming to have an income of zero decreases by 1.12 percentage points (Column 4). Consistent with the assumption that it is harder to misrepresent one's income if the registration happens during a home visit, neither the share of families with a reported income below the eligibility threshold (Column 2) nor the share of families reporting an income of zero change significantly for families who registered during a home visit (Column 5). Although a home visit allows administrators to judge a family's true income more accurately, it is conceivable that it also makes it easier to bribe the responsible social worker. Under this scenario, the absence of a significant change would not so much indicate a low level of underreporting, but rather a continued high level of underreporting even after the

34. Note that this is already not guaranteed if M is such that the highest reported income of an included family is less than $y(x^*)$.

35. Recent research suggests that this presents a bigger challenge for poorer families (e.g., Shankar et al., 2011; Mani et al., 2013), which might exacerbate the distortion.

36. The sixth wave of the World Value Survey included a question whether this is justifiable behavior. If Brazilian states were countries, the relatively rich Espírito Santo would have the strongest norms of any country against illegally claiming benefits (before the Netherlands), whereas the state of Alagoas would have the weakest norms of any country (after Mexico).

37. The entire Cadastro Único, not just a sample of families on the margin of the program.

TABLE 3.5
UNDERREPORTING DECREASES AFTER RANDOM AUDITS

	Eligible income (%)			Income R\$0.00 (%)		
	(1) Total	(2) Home	(3) CRAS	(4) Total	(5) Home	(6) CRAS
Past audit	-1.090* (0.506)	-0.219 (1.282)	-1.309* (0.543)	-1.115* (0.530)	-0.762 (0.904)	-0.997+ (0.516)
Control mean	60.505	55.952	60.512	10.823	9.779	10.981
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.946	0.742	0.940	0.844	0.673	0.844
N(municipalities)	5539	5504	5539	5539	5504	5539
N	33226	31544	33226	33226	31544	33226
Years	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017

Notes. This table reports the effect of the random audits on the distribution of self-reported income in the Cadastro Único. The dependent variable in Columns (1) to (3) is the percentage of families in a municipality who report an income that made them eligible at the time of registration. The dependent variable in Columns (4) to (6) is the percentage of families in a municipality who report having zero income. Columns (1) and (4) present the results for all families in the municipality, Columns (2) and (5) for families that were registered during a home visit, and Columns (3) and (6) for families that registered at the CRAS. “Past audit” indicates that a municipality has been audited at random. All models include municipality and year fixed effects. Standard errors are clustered at the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

audits. This interpretation, however, is unlikely as reporting an income of zero and reporting a qualifying income are both significantly less common for families registered at home ($P = 0.000$, two-sided t -tests).

In the case of Sete Quedas, CGU auditors cross-referenced data on formal employment and pensions with data from the Cadastro Único and inspected the homes of 31 families with inconsistent records. The per capita income of seven of these families made them ineligible for Bolsa Família. Three of these families reported an income of zero, while the others reported incomes that are close to the extreme poverty or the eligibility threshold at the time of their first registration. Further investigations revealed that most families deliberately “forgot” to mention a source of income or to register a family member who receives a pension. The families have subsequently been excluded from the program.

This anecdote suggests an additional test of whether income underreporting and mistargeting decrease after a municipality has been audited at random. Immediately after the audit, exclusions from Bolsa Família should increase, before dropping to a lower level than before the audit, as municipalities inspect self-reported income more closely and are less likely to admit ineligible families. As predicted, Figure A3.14 in Appendix A3.V shows that exclusions from the program increase immediately after a municipality has been audited at random, before dropping to a lower rate than before the audit ($P = 0.106$ and $P = 0.047$, respectively; see Table A3.20 in Appendix A3.VI). After a random audit, municipalities are also somewhat more likely to conduct home visits as part of the registration process, although the increase is not statistically significant ($P = 0.156$, two-sided t -test).

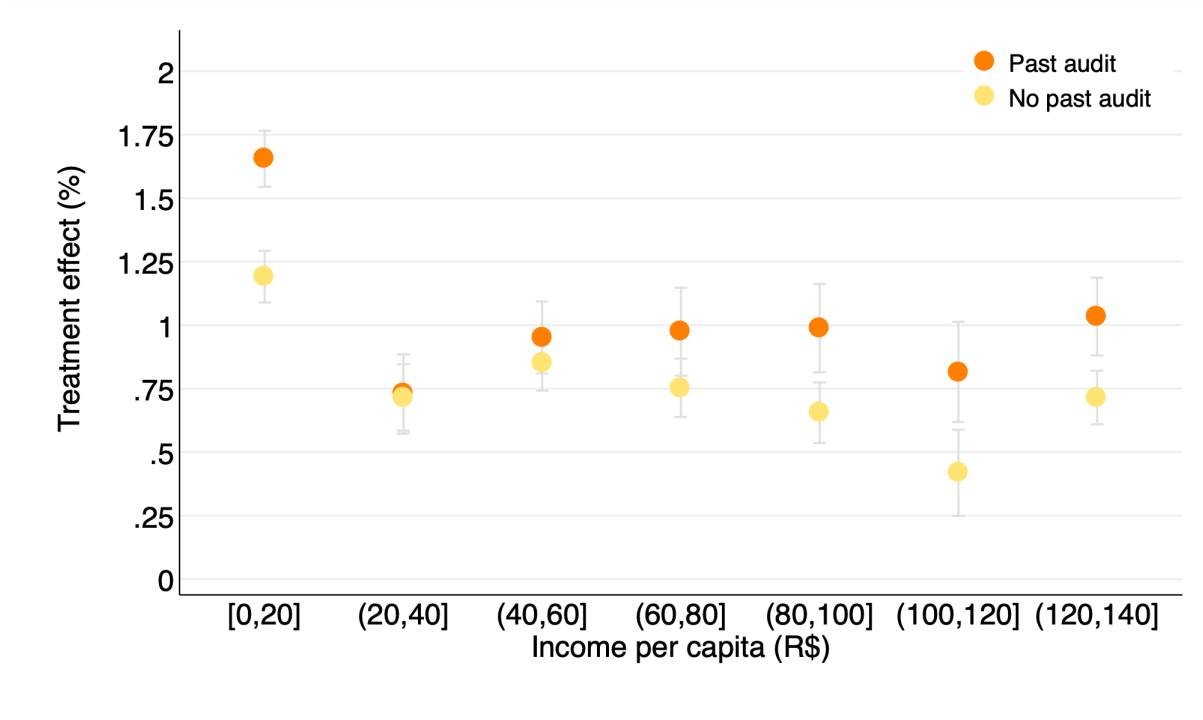


FIGURE 3.5

Program Effectiveness Increases Most for the Lowest and Highest Self-Reported Incomes

Notes. This figure displays the estimated effect of inclusion in the Bolsa Família program on children's school enrollment by income brackets. Treatment effects were jointly estimated in the most representative sample by interacting the treatment indicators in Equation (3.4) with an indicator for the family's income bracket. Error bars indicate standard errors of the estimated treatment effect and are clustered at both the family and municipality level.

The second part of the model illustrates how income underreporting translates into lower expected treatment effects. If the effectiveness gains are driven by improvements in the targeting of low-income families and if it is easier for families to underreport their income if no home visit is conducted as part of the registration process, the effectiveness gains should be concentrated among families that registered at the registration center (CRAS), where income verification is more effortful. In families that registered at home, Bolsa Família is expected to be equally effective at incentivizing school enrollment irrespective of the audits. Table A3.19 in Appendix A3.VI shows that this is indeed the case: while the treatment effect increases by at least 25% for families registered at the registration center ($P = 0.025$, $P = 0.037$, and $P = 0.065$, for the three samples), there is no change for families who registered during a home visit—point estimates for the interaction are much smaller and not statistically significant.

The families that underreported their income in Sete Quedas fell in two categories: families that report an income of zero and families that report close to the eligibility threshold. The theoretical model indeed predicts that families with incomes close to zero and families with relatively high incomes have the strongest incentives to underreport. Figure 3.5 shows that, in line with these predictions, the effectiveness gains after a random audit are concentrated at the

lower and the upper end of the income distribution.³⁸ In unaudited municipalities, families in the lowest income bracket are only 1.19 percentage points more likely to send their children to school once they are included in Bolsa Família. In contrast, after a municipality has been audited at random, Bolsa Família increases school enrollment by 1.66 percentage points, a relative gain of almost 40% ($P = 0.004$). While there are no significant gains for families in the next three income brackets, treatment effects increase by 50% ($P = 0.095$), 95% ($P = 0.082$), and 45% ($P = 0.084$), respectively, for the three highest income brackets. Note that this is compatible with increased income underreporting close to the eligibility cut-off, as the income threshold was less than R\$ 100 when the earliest families in the sample registered and has since been increased several times.

Thus, the income underreporting model is consistent with the observed patterns in the administrative data: immediately after a random audit ineligible families are excluded from the program and, going forward, fewer families report an eligible income or an income of zero. As a result, the Bolsa Família can be more precisely targeted, and its effectiveness increases, especially for the income levels with the highest predicted misreporting.

D. Testing the Model in a Field Experiment

To see if local administrators are indeed less likely to register ineligible families in municipalities that have been audited at random, I conducted a field experiment with 6,998 Bolsa Família registration centers (CRAS).³⁹ Registration centers were contacted asking about the possibility of receiving Bolsa Família and the information provided in the message was experimentally varied to make the sender eligible or ineligible while holding other characteristics constant. Consistent with the income underreporting explanation, centers in audited municipalities differentiate more between eligible and ineligible families: they are less likely to engage with ineligible families and to incorrectly state that a sender's income is compatible with Bolsa Família. For a more detailed description and additional results of the field experiment, see Appendix A3.III.

Over several months, three emails were sent to registration centers that provided an email address as part of their official contact details. The emails asked about registering for the Bolsa Família program and provided information that makes the sender either eligible—per capita income < R\$ 170—or ineligible (see Table 3.6). Relative to the "Ineligible" treatment, "Eligible I" varied the number of children and "Eligible II" reduced the reported income.⁴⁰

38. Note also that while Bolsa Família increases school enrollment at all income levels, the poorest families gain most from the program, in line with the assumptions of the model.

39. The experiment was approved by the Human Subjects Committee of the Faculty of Economics, Business Administration, and Information Technology at the University of Zurich (*OEC IRB # 2019-010*) and was preregistered at the AEA RCT registry under the number *AEARCTR-0004151*.

40. Emails were sent in three waves at the beginning of May, June, and July 2019. Within waves, emails were sent at a random time on a workday between 9:00 and 17:00 in the centers time zone. The order of the emails and their timing was randomized at the municipality level and block-randomized with respect to states and whether a municipality has been audited at random. Roughly a quarter of emails could not be delivered

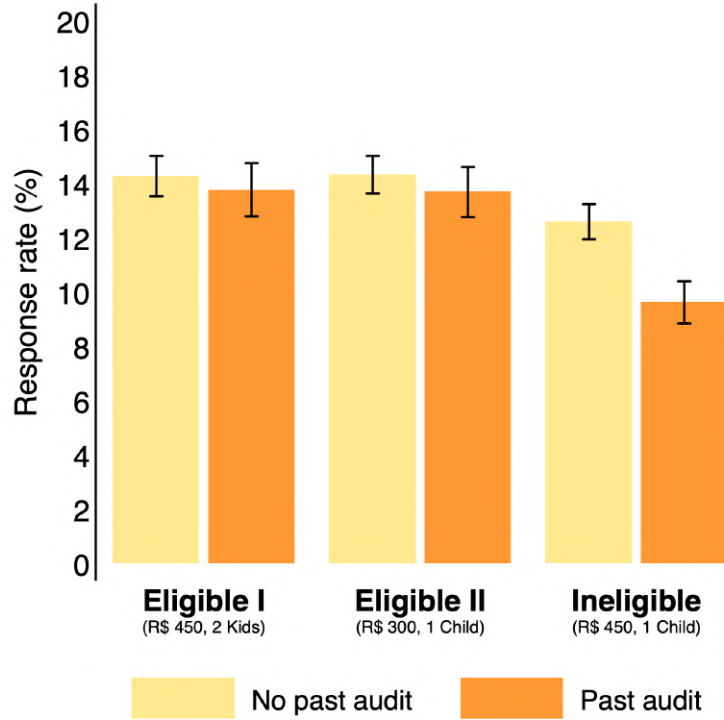


FIGURE 3.6

Response Rate to Requests from Eligible and Ineligible Families

Notes. This figure displays the difference in response rates in the three experimental conditions of the field experiment. Error bars indicate standard errors and are clustered at the municipality level.

TABLE 3.6
EMAIL TEXT BY TREATMENT

Eligible I	Eligible II	Ineligible
Greetings,	Greetings,	Greetings,
My family recently moved here, and I would like to register for Bolsa Família. I make around R\$ 450 a month and I live alone with my two children. Can you help me to register?	My family recently moved here, and I would like to register for Bolsa Família. I make around R\$ 300 a month and I live alone with my child. Can you help me to register?	My family recently moved here, and I would like to register for Bolsa Família. I make around R\$ 450 a month and I live alone with my child. Can you help me to register?
Thank you in advance.	Thank you in advance.	Thank you in advance.

Notes. This table displays the three experimental conditions in the field experiment.

and returned an error message from the host. This failure rate is independent of the wave of the experiment and whether a municipality has previously been audited or not ($\chi^2(5) = 3.248$, $P = 0.662$; see also Figure A3.2 in Appendix A3.III). For most of the analysis, these messages are excluded. However, Table A3.11 in Appendix A3.III shows that treating delivery errors as non-responses does not alter the results.

Figure 3.6 shows that registration centers were significantly more likely to reply to requests from eligible families. Consistent with the hypothesis that local corruption makes it easier for ineligible families to gain access to the program, the effect is significantly larger for centers in municipalities that have previously been audited. In unaudited municipalities, requests from ineligible families were less likely to receive a response than requests from eligible families—1.68 percentage points relative to the Eligible I treatment ($P = 0.012$) families and 1.73 percentage points relative to the Eligible II treatment ($P = 0.010$). The differences increase to 4.15 percentage points ($P = 0.000$) and 4.07 percentage points ($P = 0.000$), respectively, if a municipality has been audited at random. The treatment effects are significantly stronger in audited municipalities ($P = 0.024$ and $P = 0.031$ for the comparisons with Eligible I and II conditions, respectively). Because the response rates are relatively low, these numbers imply that requests about registration with ineligible details were approximately 12% less likely to receive a response in unaudited municipalities, but roughly 30% less likely to receive a response in municipalities that have previously been audited.⁴¹

Because municipalities are randomly selected for audits within states, I pre-registered that I would run the regression with state fixed effects to account for the stratification. Table 3.7 shows that the effect persists if state fixed effects are used (Column 1), if only within registration center variance is exploited (Columns 2 and 3), and if the control variables from Avis et al. (2018) are used, as specified in the pre-analysis plan (Columns 4 and 5). In addition, Columns (3) and (5) control for extensive design fixed effects: the order of emails, the different subject lines, the day of the week and the exact time of day the emails were sent. Table A3.13 in Appendix A3.III shows that the effects persist if the two control treatments are included separately in the regressions.

As each center receives three similar emails, responses to emails in later waves might be affected by the emails in earlier waves: the email might look familiar to social workers or be more likely to end up in a spam filter.⁴² To address these concerns, I pre-registered a robustness check to show that the effect persists if only the first wave of emails is used. The effect is even more pronounced in the first wave of the experiment. Request from ineligible families are significantly less likely to receive a response in municipalities that have been randomly audited in the past—7.35 percentage points compared to the Eligible I treatment ($P = 0.000$) and 6.36 percentage points compared to the Eligible II treatment ($P = 0.001$). In unaudited municipalities, the effects are much weaker—2.78 percentage points compared to the Eligible I treatment ($P = 0.061$) and 0.91 percentage points compared to the Eligible II treatment ($P = 0.524$). The treatment effects are significantly stronger in audited municipalities ($P = 0.049$ and $P = 0.016$ for the comparisons with Eligible I and II, respectively). See also Table A3.12 in Appendix A3.III.

41. For a discussion of why the response rates are relatively low, see Appendix A3.III.

42. The response messages were screened for signs of suspicion. Overall, fewer than 1% of the responses showed any sign of suspicion. However, this rate increases with each round from 0.27% in the first round to 1.52% in the final round.

TABLE 3.7
RESPONSE RATES TO REQUESTS FROM ELIGIBLE AND INELIGIBLE
FAMILIES

	(1)	(2)	(3)	(4)	(5)
Ineligible	-1.708** (0.589)	-1.526* (0.597)	-1.500* (0.602)	-1.672** (0.589)	-1.686** (0.591)
Past audit	0.455 (1.000)			0.250 (0.932)	0.330 (0.932)
Ineligible \times Past audit	-2.396* (0.932)	-2.703** (0.938)	-2.694** (0.944)	-2.435** (0.932)	-2.512** (0.942)
Population (Log.)				3.510*** (0.488)	3.509*** (0.489)
Income inequality (Gini)				-11.855+ (6.139)	-12.170* (6.167)
Income per capita (Log.)				5.632** (1.764)	5.618** (1.766)
Illiteracy				0.015 (0.075)	0.020 (0.075)
Urban population				-0.707 (2.234)	-0.532 (2.244)
Control mean	14.303	14.371	14.371	14.303	14.303
State FE	Yes	Yes	Yes	Yes	Yes
Center FE	No	Yes	Yes	No	No
Order FE	No	No	Yes	No	Yes
Subject line FE	No	No	Yes	No	Yes
Day FE	No	No	Yes	No	Yes
Time FE	No	No	Yes	No	Yes
R2	0.039	0.599	0.606	0.068	0.078
N	15891	15736	15736	15891	15891

Notes. This table reports the difference in response rates to requests from eligible and ineligible families in the field experiment. The dependent variable is an indicator that takes value 100 if the registration center replied to a request and 0 otherwise. “Ineligible” indicates that the details in the request made a family ineligible for Bolsa Família, “Past audit” indicates that a municipality has been audited at random, and “Ineligible \times Past audit” is the interaction of the two treatments. “Population (Log.)”, “Income inequality (Gini)”, “Income per capita (Log.)”, and the rates of “Illiteracy” and “Urban population” control for municipality characteristics in 2000, before the inception of the audits program. Columns (2) and (3) include registration center fixed effects. Columns (3) and (5) include fixed effects for the order of emails, the different subject lines, the day of the week and the exact time of day the email was sent. All models include state fixed effects. Standard errors are clustered at the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Even if the response rates in different experimental conditions were the same, employees of registration centers might still discern between eligible and ineligible families when they compose their reply. Thus, in addition to testing for differences in the response rate, each response was also coded to analyze the content of the message. Table A3.10 in Appendix A3.III test for differences in the content of the emails, using a two-step Heckman selection model to control for the different response rates in the experimental conditions and in previously audited municipalities. Responses in the Ineligible treatment were 6.74 percentage points more likely to contain an incorrect assessment of the family’s eligibility, i.e., to state that the family qualifies for

Bolsa Família ($P = 0.024$). However, this effect is driven exclusively by unaudited municipalities; it is fully offset by the negative 7.41 percentage point interaction term ($P = 0.027$), suggesting that audited municipalities indeed pay more attention to families' incomes. Messages do not significantly differ in how much practical information they include, nor is there a difference in whether they explain the eligibility criteria of Bolsa Família. Finally, emails were also screened for direct offers of collusion and for less overt signals of corruption, such as suggesting that the rules are flexible. However, none of the messages contained such a smoking gun.

Although responding to emailed requests is an imperfect proxy for the likelihood of including a family in the Cadastro Único, the results of the field experiment are consistent with the income underreporting mechanism: randomly audited municipalities pay closer attention to families' eligibility.

E. Social Norms on Income Underreporting

The lower rates of income underreporting after a random audit might reflect changes in social norms, rather than differences in the difficulty of successfully misrepresenting one's income. For example, the experience of being audited and the revelation of the irregularities in Bolsa Família might change citizens' social norms about underreporting their income, condoning public corruption, or reporting suspected fraud. To test this alternative hypothesis, I elicited relevant social norms in an incentivized online experiment⁴³ with 675 participants living in 424 municipalities, some of which had been randomly selected for audits in the past.⁴⁴ However, there is no evidence that social norms change as a result of a random audit. For a more detailed description and additional findings of the online experiment, see Appendix A3.IV.

In the experiment, I presented participants with three short vignettes (see Table 3.8): First, a family that underreports its income to qualify for Bolsa Família. Second, a local administrator who suspects that the family underreports their income but turns a blind eye. Finally, a neighbor who calls the local registration center to blow the whistle on the family. Participants are then asked to rate the behavior in the scenario, given four choices: very wrong, somewhat wrong, somewhat right, very right. To avoid that participants give socially desirable but untruthful responses, I used the method developed by Krupka and Weber (2013): instead of asking participants for their opinion, they are incentivized to try and give the same response as another randomly selected participant. At the end of the experiment, one of the three vignettes was randomly selected for payment, and participants received R\$ 10 if the responses matched. This

43. The experiment was approved by the Human Subjects Committee of the Faculty of Economics, Business Administration, and Information Technology at the University of Zurich (*OEC IRB # 2019-008*).

44. Participants were recruited through Facebook to achieve maximum geographic coverage while maintaining relatively precise targeting. Participation was restricted to resemble the typical Bolsa Família beneficiary: women aged 18 to 50, who have an interest in Bolsa Família and are accessing Facebook on their mobile device. The 31 municipalities that are too populous for the random audits were excluded. Summary statistics for participants are displayed in Table A3.14 in Appendix A3.IV.

TABLE 3.8
VIGNETTES AND ELICITED SOCIAL NORMS

Underreporting one's income	Turning a blind eye	Blowing the whistle
The family of Sônia would like to receive money through the Bolsa Família program. She knows that their income is too high to qualify, so she reports a lower income when she registers in the Cadastro Único.	The employee at the CRAS suspects that Sônia's income is higher than what she reports. He makes a deal with Sônia so that she can nevertheless benefit from the Bolsa Família program.	Now suppose that her neighbor knows that the family of Sônia makes too much money to qualify for Bolsa Família. The neighbor decides to report it to the social council.

Notes. This table displays the three vignettes used in the norm elicitation game.

transforms the question into a coordination game in which the social norm serves as a focal point.

Figure 3.7 demonstrates that social norms do not significantly change after a random audit. The distribution of responses is almost identical in previously audited and unaudited municipalities for all three vignettes ($\chi^2(3) = 0.162$, $P = 0.984$; $\chi^2(3) = 1.543$, $P = 0.672$; $\chi^2(3) = 0.406$, $P = 0.939$, χ^2 -tests). Using ordered logistic regression, Table A3.15 in Appendix A3.IV shows that the participants from audited municipalities are just as accepting of these behaviors, and that this result is robust to controlling for state fixed effects, individual participant characteristics, and the usual controls for municipalities' socioeconomic development.

Additional tests in Appendix A3.IV show that there is also no shift in participants' beliefs about the prevalence of these behaviors (Table A3.16) and that there is no change in general rule-following, measured by an incentivized Fischbacher and Föllmi-Heusi (2013) honesty game (Figure A3.5). Moreover, there are only minor differences in municipality and participant characteristics between audited and unaudited municipalities in the online sample (Table A3.14).

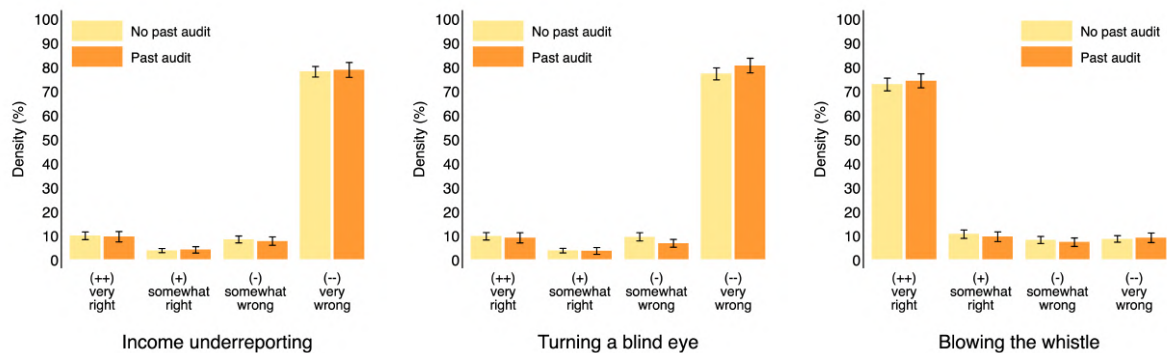


FIGURE 3.7
Social Norms Don't Change after a Random Audit

Notes. This figure displays the distribution of responses in the Krupka and Weber (2013) norm elicitation game. Error bars indicate standard errors and are clustered at the municipality level.

Thus, there is no evidence that changes in social norms can explain the lower levels of income underreporting after a municipality has been audited at random.

VI. FINANCIAL GAINS FROM BOLSA FAMÍLIA

In this paper, I argue that Bolsa Família has reasonably effective safe guards against clientelism and embezzlement. This claim is based primarily on the program's design—payments directly to beneficiaries' cards instead of bulk payments to the local administration and an anonymized central process to select beneficiaries. While this prevents outright embezzlement of funds and the trading of program access for bribes or votes, it does not completely rule out the possibility of financial gains from the program.

If local officials want to benefit financially from the program, they are left with three options: retaining benefit cards, registering themselves in the Cadastro Único, and diverting other funds connected to local registration centers or the complementary programs municipalities are required to offer for Bolsa Família recipients. In this section, I show that while each of these practices has been known to happen, they cannot account for the effectiveness gains of Bolsa Família after a municipality has been audited at random.

A. *Retaining Benefit Cards*

Retaining benefit cards allows a corrupt official to withdraw the payments from the account of legitimate beneficiary families. This was a major concern in the early years of the program but has since become very uncommon.

Once a family is admitted to the Bolsa Família program, the Caixa provides it with a magnetic stripe card that can be used to withdraw the benefits at branches of the Caixa, special ATMs, lottery points, postal offices, and certain shops. Most cards are delivered through the postal system and beneficiaries need to sign a receipt. If this is not possible, the cards are returned to the Caixa that then attempts to deliver them through other channels, e.g., by inviting families to pick them up at the nearest Caixa branch or by sending them to another location specified by the beneficiary. Once families have received their card, they have to register their PIN number at the nearest Caixa branch before they can access any benefits.

There are several ways a corrupt official can take possession of a families card. Prior to 2006, local branches of the Caixa were responsible for distributing the cards to beneficiary families and there were several incidents where cards were not immediately delivered to beneficiaries or where there was no proof of receipt (Lindert et al., 2007). Now that most cards are delivered through the postal system, fewer cards are vulnerable to this kind of fraud. Alternatively, a corrupt administrator can seize the card under a pretext or a store owner can refuse to return the card to a customer to force her to spend the benefits at his store. Finally, in remote areas,

some people will offer to go to town and collect the payments for multiple families for a cut of the benefits.

The MDS receives 10-20 reports of retained cards per year—relatively few considering the almost 14 million benefit cards in circulation. Sete Quedas is representative in that the MDS has never received a complaint about retained cards since they started to collect this information in 2010; in an average year, the MDS receives not a single complaint from more than 99.7% of municipalities. As the financial damages incurred through this form of corruption are roughly proportional to the number of retained cards, the total losses are relatively minor. Even if we assume that only one in a hundred affected cards is being reported, 99.99% of payments would still be unaffected. Thus, funds stolen by retaining benefit cards cannot account for the effectiveness gains of Bolsa Família.

B. Self-Registration in the Cadastro Único

"Beneficiary families with a member employed by the city hall underreported their income." (CGU, 2015)

Like any other citizen, corrupt officials can benefit by registering themselves or a family member in the Cadastro Único and providing inaccurate information that increases the probability of being included. Unlike most families, however, local officials and program administrators can potentially exert control over the registration and income verification process. Because of this, the MDS and the CGU periodically cross-check beneficiaries with databases on elected officials and public employees. Billy the cat, who made a brief appearance in Section II, actually belonged to a local program administrator (Hider, 2014).

The financial impact of this strategy, however, is relatively small. The fraction of payments syphoned to public employees can be quantified based on audits where CGU auditors cross-referenced public employment records with data from the Cadastro Único. In the case of Sete Quedas, for example, CGU auditors discovered one Bolsa Família beneficiary who had been employed by the municipality for over a decade and earned a monthly wage of more than R\$ 2,100. Despite some news-worthy cases, less than 0.75% of payments are affected in any given year.⁴⁵

To put this number into perspective, we can compare it to the average share of families uncovered to have underreported their income in the random samples of families investigated by the CGU (see Figure A3.12 in Appendix A3.V). This share varies from 8.5% (in 2014) to 23.9% (in 2010). This is in line with the findings from similar programs in Indonesia (Alatas et al., 2019) where local elites had significantly more control over the allocation of benefits; while

45. The number of cases was manually extracted from the CGU reports by a Brazilian research assistant and it covers municipal, state, and federal employees. The estimate is conservative, as it also includes cases where public employees were entitled to some, but not all of the benefits they received.

local elites capture some of the benefits, the effect is not economically large and tends to pale in comparison to the costs of other targeting errors.

C. Embezzling Complementary Funds

Diverting other funds from local registration centers can potentially be quite lucrative compared to the other two strategies. If corrupt officials embezzle funds that are designated for the local registration centers, the resulting infrastructure and staff shortages might affect the performance of the Bolsa Família program. Moreover, municipalities are required to offer complementary programs designed to help Bolsa Família recipients to comply with the program's conditionalities and to realize lasting improvements in the standard of living. Here too, significant amounts can go missing. Section VII discusses these points in more detail and shows that changes in infrastructure, registration center employment, and the availability of complementary programs cannot account for the gains in program effectiveness.

VII. ALTERNATIVE EXPLANATIONS

Until now, the analysis has focused primarily on income underreporting and the resulting mistargeting as an explanation for the effectiveness gains of Bolsa Família after a random audit. In this section, I show that while common findings from the CGU's audit reports provide anecdotal evidence for several additional mechanisms how random audits might affect the effectiveness of the Bolsa Família program, none of these candidates—closer school attendance monitoring, higher data quality, changes in administrative processes, better infrastructure and funding, complementary social programs, tighter governance, or increased whistleblowing—can explain the increase in Bolsa Família's effectiveness.

A. School Monitoring

"School attendance records of students benefiting from the Bolsa Família program entered in the Projeto Presença system by the municipality's program manager are in disagreement with those found in class books." (CGU, 2015)

Municipalities' other major responsibility, besides registering families, is collecting the data to monitor compliance with Bolsa Família's conditionalities—most importantly children's school attendance. However, there is no evidence that school attendance records become more accurate after a random audit.

Bolsa Família depends on municipal administrators to report if children of beneficiaries fail to attend school regularly enough to comply with the program's conditionalities. Teachers record students' absences in the class book, which is in turn used by the school administrators to report attendance in the ministry of education's Projeto Presença system. As the auditors

explain, insufficient monitoring risks that Bolsa Família provides only short-term relief but no sustainable progress in the fight against poverty and social marginalization (CGU, 2015). Thus, Bolsa Família will be less successful if corrupt administrators don't fulfill their responsibility or collude with families to overstate compliance with conditionalities.

In the case of Sete Quedas, several students were given 99% attendance scores in the monitoring system despite having insufficient attendance rates for the two months scrutinized during the audit. Sete Quedas' negligence to adequately monitor school attendance is relatively benign compared to the failures uncovered in some other municipalities, where "class books are not traceable", "students who benefit from Bolsa Família cannot be located", and school officials "lack knowledge of their responsibilities", "report on the attendance of students enrolled in other schools", or "report 100% attendance for all students without any documentation."

Unfortunately, the Cadastro Único does not contain information on students' school attendance rates. However, the audit reports can be used to see whether school monitoring improves in municipalities that have been audited at random. Using the specification in Equation (3.3), there is no significant change in the number of irregularities related to attendance monitoring in municipalities that have previously been audited ($P = 0.960$). Additionally, temporary blockage of benefits due to non-compliance with the conditionalities of Bolsa Família, most importantly school attendance, can be used as a proxy for how closely families are monitored. Table A3.20 shows that even though the number of families with temporarily blocked benefits increases significantly immediately after a random audit ($P = 0.032$), this is not a lasting change and there is no significant change in the long run ($P = 0.811$). Thus, improvements in school monitoring are unlikely to explain the lasting effectiveness gains of Bolsa Família after a municipality has been audited at random.

B. Data Inconsistencies

"[...] the local manager should update the registry entries of the beneficiaries indicated in the inspection report, to adjust the data recorded in the CadÚnico with the real family structure." (CGU, 2015)

Mistargeting can originate not only from deliberately misleading statements during the registration process but also from unintentional errors and outdated information in the database. Because the MDS relies on local administrators for the registration and monitoring of beneficiaries, it reimburses municipalities with a payment per household per month if the municipality fulfills its responsibilities sufficiently well.

To assess municipalities' implementation of Bolsa Família, the MDS constructs a monthly index of municipal management quality (IGD-M) based on the consistency of data in the Cadastro Único and the school and health monitoring systems. The index is a weighted sum of the rates of school monitoring (the fraction of children in beneficiary families with updated entries

in the Projeto Presença system), health monitoring (the fraction of families subject to medical conditionalities that is covered in the health monitoring system), and the fraction of up-to-date entries in the Cadastro Único. Until July 2015, the estimated coverage rate of the Cadastro Único also contributed to the index.⁴⁶ If the IGD-M and its subindices exceed a certain threshold and some additional administrative requirements are satisfied, municipalities receive IGD-M \times R\$ 3.25 per valid registration, plus additional incentive payments, for example, to follow up on families with suspended benefits.

If Bolsa Família becomes more effective because municipalities improve their data management, this will be reflected in the IGD-M and its components. Table A3.21 in Appendix A3.VI shows that this not the case. The index is unchanged, as are the subindices for health, data updating, and the coverage rate of the registry. If anything, school monitoring worsens slightly. The absence of changes in these indicators, however, conceals significant changes in both the nominators and the denominators that make up these fractions: There are fewer children whose school attendance is monitored, but also fewer children in beneficiary families. There are fewer families whose medical checkups are monitored, but also fewer families required to do the check-ups (see Table A3.22 in Appendix A3.VI). In short, significantly fewer families are registered as being eligible, again suggesting that municipalities exclude ineligible families in the aftermath of an audit and are more careful going forward.

C. CRAS Processes

"[...] the identified problems are caused by the lack of pre-established and properly formalized routines for verifying and monitoring compliance with the legislation that governs the program." (CGU, 2015)

When asked to explain irregularities in the income of registered families, municipal administrators in Sete Quedas resorted to case-by-case explanations and tried to place the blame solely on the families. According to the auditors, they did not sufficiently appreciate that these cases are symptomatic of inadequate processes at the local registration center (CGU, 2015).

The IGD-M is a rather crude tool to monitor the registration centers and it is mostly used to incentivize consistency between the interlinking computer systems. More informative about the actual processes and practices is the annual census of social assistance, Censo SUAS, that is completed by the social assistance centers (CRAS). The survey collects basic information on employees and detailed information on physical infrastructure, technical equipment, and administrative processes—including how the Cadastro Único is updated. Questions include whether the CRAS updates the registry,⁴⁷ which employees work with the database, and more

46. Later, the coverage rate was so close to 100% that the MDS stopped using this information. See Figure A3.6 in Appendix A3.V.

47. Not all centers input data in the Cadastro Único because some lack the necessary infrastructure and training or focus on other forms of social assistance.

recently also in what format data is initially collected.

The results in Table A3.23 in Appendix A3.VI suggest that the auditors' well-meaning advice is largely ignored: centers continue to update the Cadastro Único in pretty much the same way as before the audit. Centers are equally likely to update the Cadastro Único, just as likely to have a special team to work with the registry, and individual employees are equally likely to handle the database. When collecting information on families, CRAS centers that do update the Cadastro Único stick to the method they have always used and are just as likely to rely on paper or to enter the data digitally than before the audit.

Table A3.24 in Appendix A3.VI fails to find any significant change in the composition of the 80,000-strong workforce at the CRAS. Employees have the same age and gender profile, the same experience, similar working hours, and the same educational level: the shares of employees with at most completed primary education, secondary education, some college, a college degree, or a post-graduate qualification all remain unchanged. There is also no change in the legal aspects of the employment relationships: employees are just as likely to be hired under the rigid Consolidation of Labor Laws (CLT) after an audit as before, and the same holds for civil servant appointed by the mayor under the discretionary rules of Art. 37.

Finally, CRAS centers in audited municipalities are also not more likely to seek citizen participation to ensure that their activities suit the needs of their clients. Table A3.25 in Appendix A3.VI shows that if anything, centers are less likely to invite citizen participation to improve the center's services: the share of centers that report not soliciting any input from the population—formal or otherwise—increases by 7.59 percentage points, albeit not significantly ($P = 0.163$). Moreover, centers are 2.51 percentage points less likely to have elected citizen representatives ($P = 0.076$) and 5.44 percentage points less likely to have a citizen committee after a random audit ($P = 0.095$), making it unlikely, that increased citizen participation is behind the observed performance gains.

Overall, changes in practices at the CRAS are insufficient to account for the improved performance of Bolsa Família after a municipality has been audited at random.

D. CRAS Infrastructure and Funding

"Used and broken toys were delivered to the Tia Solíria Day Care Center." (CGU, 2015)

One way by which local corruption can hinder the effectiveness of government policy is through its effect on municipal infrastructure: even if the money from Bolsa Família payments reaches beneficiaries, the program could be less effective if the CRAS centers lack the infrastructure to properly serve families in the municipality.

In the previous rounds of the audit program, the infrastructure at the CRAS centers in

several municipalities was inspected and found lacking.⁴⁸ Unfortunately, the auditors did not inspect the infrastructure at the CRAS in Sete Quedas, but the mechanism is well-illustrated by a curious incident involving a woman the auditors identify as "the first-lady of Sete Quedas" (CGU, 2015): About half a year before the audit, the municipality spend R\$ 4,614.78 it had received from the National Fund for Education Development to buy educational toys and sporting equipment for the Tia Solíria Day Care Center. At the time of the audit, the toys were in the nursery, but they were in two garbage bags—unpacked, dirty, broken, and incomplete. Asked about the garbage bags, the head of the nursery said that they had been delivered in the previous week by the first lady of the municipality. Having initially given the toys to other unidentified families, she scrambled to retrieve them when she learned of the impending audit.

However, as shown in Figure A3.15 in Appendix A3.V, major changes in center infrastructure are as conspicuously absent as changes in the workforce and practices of registration centers: For the 37 analyzed survey items in the Censo SUAS—covering everything from the physical infrastructure (number and type of rooms), whether they comply with the norms for accessibility in public buildings (ABNT NBR9095), the ownership of the premises, IT and other technological infrastructure, vehicles available to the CRAS staff, to whether the center has a toys and sporting equipment at its disposal—, there is not a single significant change after correcting for multiple hypothesis testing (FDR; Benjamini and Hochberg, 1995).⁴⁹

Similarly, one might wonder whether municipalities allocate more funds for the proper administration of their social programs. However, Tables A3.26 and A3.27 in Appendix A3.VI show that neither the amount municipalities spend on social programs and their administration nor the amount they spend on education increases significantly after a municipality has been audited at random. This result holds irrespective whether expenditure is analyzed in absolute numbers, per capita, or as a share of the total municipal budget. If anything, municipalities decrease per capita social expenditure by about 4.1% ($P = 0.061$). Thus, neither changes in the infrastructure nor the funding for social programs can account for the increased effectiveness after a random audit.

E. Complementary Actions and Programs

"[...] it was verified that the Municipality did not offer complementary programs to Bolsa Família." (CGU, 2015)

Complementary programs for beneficiary families—literacy classes, occupational training, microcredits, and guidance in accessing government services—are an important ingredient of Bolsa Família's strategy to overcome poverty in a sustainable way (CGU, 2015). Municipalities

48. Common findings include lack of computers (e.g., CGU, 2014a), problems with of accessibility (e.g., CGU, 2014b), and absence of sufficiently large rooms for communal activities (e.g., CGU, 2014c).

49. Without multiple hypothesis correction, only one item changes significantly at the 5% level: after an audit, centers are more likely to occupy a building owned by the municipality ($P = 0.035$).

play a key role in the development of these services and are especially called upon to assist families that are in breach of their Bolsa Família conditionalities. It is straightforward to see how improvements in these programs after a random audit might account for the effectiveness gains that we observe. This explanation, however, is not supported by the data; there is no significant change in complementary programs after a random audit.

While the Census SUAS does not explicitly ask about complementary programs for Bolsa Família, it does ask about activities and programs aimed at vulnerable families in general through the Serviço de Proteção e Atendimento Integral à Família (PAIF). Although PAIF does not exclusively serve families in Bolsa Família, a large number of its activities focus on low-income families (Afonso et al., 2013) so that PAIF activities are a good proxy for the existence of complementary programs.

Figure A3.16 in Appendix A3.V shows the effect of a random audit on 23 programs and activities targeted at vulnerable families. None of the changes are significant at the 5% level, even without correction for multiple hypothesis testing (Benjamini and Hochberg, 1995). Centers don't change their outreach programs to welcome new families, don't offer different programs and activities, don't provide more specialized coaching for families in various life situations, and are just as likely to refer families to other social and public services than before the municipality has been audited. Most relevantly for Bolsa Família, there is no evidence that CRAS centers are more likely to coach families that are in breach of their conditionalities ($P = 0.909$), nor that they are more likely to help families to register or update their data in the Cadastro Único ($P = 0.950$).

F. Governance and Oversight

"The Municipal Council of Social Assistance does not fulfill its obligations to monitor and inspect the programs and services." (CGU, 2015)

Brazilian municipalities are required to establish a Council for Social Assistance (CMAS) to monitor the local provision of social services. More conscientious governance could account for the effectiveness gains of Bolsa Família if social councils monitor programs more closely after a random audit. This explanation, however, is not consistent with the data from the Censo SUAS.

Although the social council's powers and responsibilities vary from place to place, they usually involve approving plans and budgets for local social services, establishing the rules to grant special relieve to families hit by certain life-events, and monitoring and overseeing the various social and welfare programs. Social councils should be composed of representatives of the municipal administration and other organizations providing social assistance, as well an equal number of members from civic organizations, program beneficiaries, and other representatives of the public. In many cases, the regulations specify that the presidency rotates between

representatives of the administration and the public.

In the case of Sete Quedas, the social council is formally responsible for monitoring all aspects of Bolsa Família—the registration and data management in the Cadastro Único, the monitoring of compliance, the temporary blockages of benefits, and the development of complementary programs.⁵⁰ When asked by the auditors, however, the members of the social council in Sete Quedas confirmed not only that they had not carried out any inspections of the municipality's social programs in the previous year but that they had not even met during this time (CGU, 2015).

Figure A3.17 in Appendix A3.V shows that social councils are unlikely to account for the increased effectiveness of Bolsa Família after a random audit: social councils are equally likely to be responsible for the program and its monitoring, to discuss results from inspections in meetings, to include a beneficiary representative, to have a commission specifically for Bolsa Família, or to receive and discuss alleged misconduct and abuse of the program. More generally, councils discuss the same topics, engage in the same activities, and are governed by the same rules. Among 53 items in the annual census of social councils, there is not a single one that changes significantly after correcting for multiple hypothesis testing (Benjamini and Hochberg, 1995).⁵¹

G. Social Control

"Civic participation in the control of the Bolsa Família Program is restricted due to non-disclosure of the list of beneficiaries of the program by the municipal administration." (CGU, 2015)

Municipalities are required to publicly post lists of all Bolsa Família beneficiaries with their names and social security numbers, similar to the disclosure made by the federal government through the Portal da Transparência. Publication of this lists, it is assumed, enables members of the public to denounce families that illegitimately claim benefits: "it should be emphasized that the disclosure of the list of Bolsa Família beneficiaries is important to make the program transparent, to identify irregularities and to allow possible denunciations by citizens" (CGU, 2015). These complaints against families and officials received through whistleblower systems are investigated as part of the audits.

In the case of Sete Quedas, the MDS did not record any denunciations of illegitimate payments from citizens or program administrators, nor any complaints against employees of the CRAS. Subsequent inspections of the public areas at the city hall and the CRAS uncovered that the municipality had failed to publish the list of beneficiaries.

50. This is increasingly the case: 73.3% of social councils had this responsibility in 2011, and the number has increased to 91.8% in 2017.

51. Without the correction, only one item is significant at the 5% level: social councils may be somewhat more likely to organize town-hall meetings ($P = 0.047$).

There is, however, no evidence that more citizens blow the whistle after a municipality has been audited at random. Table A3.28 in Appendix A3.VI shows that there is no change in the number of denunciations the MDS receives from beneficiaries, non-beneficiaries, or program administrators, nor is there a change in the number of complaints about the CRAS or its employees. Naturally, not every complaint will be made through the MDS's system and complaints with local authorities are not collected centrally. Although it cannot directly be observed whether the number of complaints made to local authorities increases significantly after a random audit, municipalities report whether they have a special ombudsman to deal with denunciations, and whether their social council received and discussed any denunciations over the last year. None of these items change significantly after a municipality has been audited at random.⁵² Thus, increased social control is unlikely to explain the improved performance of Bolsa Família after an audit.

VIII. CONCLUSION

This paper shows how corruption negatively impacts the effectiveness of government policy in a setting where bribery, clientelism, and embezzlement play at most a negligible role. Despite Bolsa Família's strong safeguards against corruption, local corruption significantly reduced its effectiveness. After a municipality has been audited at random, Bolsa Família becomes roughly 30% more effective at increasing school enrollment. Thus, even though Bolsa Família bypasses local governments for the allocation and payment of benefits to minimize the potential for corruption (Lindert et al., 2007), local corruption can still affect the effectiveness of the program if it leads to more income underreporting.

These effectiveness gains are largely driven by better targeting of the program to families that benefit the most. The mechanism is illustrated using a theoretical model of the registration process where families decide how to report their income: underreporting one's income increases the probability of being included in Bolsa Família and detection is less likely in high corruption municipalities. The model is consistent with patterns in the distribution of reported incomes and administrative data on the effects of home-visits during the registration process and the number of families excluded from the program due to income underreporting. A field experiment with registration centers provides additional evidence that income underreporting is easier in municipalities that have not been audited. In contrast, an online experiment with low-income families rules out that changes in social norms explain the shift in income reporting. Other explanations such as tampering with benefit cards, closer monitoring of school attendance, better administrative processes, improvements in infrastructure and funding for social assistance, complementary

52. See Figure A3.17 in Appendix A3.V for the result on social councils. The result on special ombudsmen is based on a regression with state fixed effects, as the question was only included in the Censo SUAS 2017, precluding the use of the more rigorous specification with municipality and time fixed effects.

programs for beneficiary families, and tighter governance and social control cannot explain the results.

Although using Bolsa Família's official database offers many advantages over secondary data sets, it limits which policy outcomes can be studied. The Cadastro Único is not a particularly rich dataset when it comes to educational outcomes and it includes little that can be used to study the program's health priorities. For example, it does not include information from the school attendance and health systems used to monitor compliance with the program's conditionalities. Using data from the ministries of health and education, future research could show whether local corruption also reduces Bolsa Família's effectiveness in promoting its health and nutritional goals and whether the school enrollment gains translate into improved test scores and better labor market outcomes.

The results of this paper speak to the positive effects of government audits (e.g., Di Tella and Schargrodsky, 2003; Olken, 2007; Bobonis et al., 2016; Avis et al., 2018): while the focus of government audits is to reduce the embezzlement of funds, they have second-order effects such as stimulating economic activity (Bologna et al., 2015; Colonnelli and Prem, 2017; Giannetti et al., 2017), improving educational attainment through better school funding (Ferraz et al., 2012) or, in this case, improved targeting of social programs. However, it is worth noting that the random audits not exclusively investigate whether municipalities can account for their use of federal funds and whether the goods and services they pay for have been delivered. As the audits look specifically for evidence of ineligible families in the Bolsa Família program, it is an open question whether the same gains would be observed if the auditors focused solely on municipalities' use of funds.

The results of the paper also have implications for the effective targeting of social programs. Bolsa Família's increased effectiveness after a municipality has been audited appear to be driven solely by the fact that the program is more likely to reach the families that respond most to the program. Borrowing from the field of personalized medicine (Kent et al., 2018), one promising approach to poverty alleviation focuses on predicting families' heterogeneous treatment responses to maximize the effectiveness of anti-poverty programs (e.g., McBride and Nichols, 2018).

Given the importance of accurate self-reporting and the challenges of income verification for the effective targeting of Bolsa Família and similar programs, future research should explore interventions that increase the likelihood of accurate reporting. Studies in other settings provide encouraging evidence of significant improvements in the honesty of self-reports from remedies as simple as having people sign at the beginning rather than at the end of a self-report (Shu et al., 2012), including a moral appeal or a reminder about the possibility of detection (Bott et al., 2017), and priming participants' religious (Randolph-Seng and Nielsen, 2007), professional (Cohn et al., 2014), and social identity (Cohn et al., 2015). Approaches could also focus on the administrator's duty to verify families' incomes, possibly inspired by interventions to curb physicians' overprescription of antibiotics (e.g., Meeker et al., 2014, 2016). This research should

focus on those design aspects of Bolsa Família that might inadvertently increase underreporting: As underreporting only increases the probability of being included, there is a moral wiggle room to think that underreporting not necessarily leads to illegitimate benefits (Dana et al., 2007). Similarly, some families who would not underreport if the social worker made the inclusion decision directly might do so if an impersonal process in the capital decides (Mazar et al., 2008).

REFERENCES

- ABDULAI, A.-G. AND S. HICKEY (2016): “The politics of development under competitive clientelism: Insights from Ghana’s education sector,” *African Affairs*, 115, 44–72.
- AFONSO, M. L. M., C. B. HENNON, T. L. CARICO, AND G. W. PETERSON (2013): “A methodological approach for working with families in SUAS: a critical reading through the lens of citizenship,” *Psicologia & Sociedade*, 25, 80–90.
- ALATAS, V., A. BANERJEE, R. HANNA, B. A. OLKEN, R. PURNAMASARI, AND M. WAI-POI (2019): “Does elite capture matter? Local elites and targeted welfare programs in Indonesia,” in *AEA Papers and Proceedings*, vol. 109, 334–39.
- ALATAS, V., A. BANERJEE, R. HANNA, B. A. OLKEN, AND J. TOBIAS (2012): “Targeting the poor: evidence from a field experiment in Indonesia,” *American Economic Review*, 102, 1206–40.
- ALDERMAN, H. (2002): “Do local officials know something we don’t? Decentralization of targeted transfers in Albania,” *Journal of public Economics*, 83, 375–404.
- AVIS, E., C. FERRAZ, AND F. FINAN (2018): “Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians,” *Journal of Political Economics*, 126, 1912–1964.
- BENJAMINI, Y. AND Y. HOCHBERG (1995): “Controlling the false discovery rate: A practical and powerful approach to multiple testing,” *Journal of the Royal Statistical Society. Series B*, 57, 289–300.
- BERTRAND, M., S. DJANKOV, R. HANNA, AND S. MULLAINATHAN (2007): “Obtaining a driver’s license in India: an experimental approach to studying corruption,” *The Quarterly Journal of Economics*, 122, 1639–1676.
- BOBONIS, G. J., L. R. CÁMARA FUERTES, AND R. SCHWABE (2016): “Monitoring corruptible politicians,” *American Economic Review*, 106, 2371–2405.
- BOHN, S. R. (2011): “Social policy and vote in Brazil: Bolsa Família and the shifts in Lula’s electoral base,” *Latin American Research Review*, 54–79.
- BOLOGNA, J., A. ROSS, ET AL. (2015): “Corruption and entrepreneurship: Evidence from a random audit program,” *Department of Economics Working Paper Series*, No. 15-05.
- BOTT, K. M., A. W. CAPPELEN, E. Ø. SØRENSEN, AND B. TUNGODDEN (2017): “You’ve got mail: A randomised field experiment on tax evasion,” *NHH Working Paper*.
- BRAZIL (2016): “Governo coloca em prática ação para barrar fraudes no Bolsa Família,” *Portal Brasil*.
- BROLLO, F., K. KAUFMANN, AND E. LA FERRARA (2019): “The political economy of program enforcement: Evidence from Brazil,” *Journal of the European Economic Association*.
- CALVERAS, A., J.-J. GANUZA, AND E. HAUKE (2004): “Wild bids. Gambling for resurrection in procurement contracts,” *Journal of Regulatory Economics*, 26, 41–68.

- CARAM, B. (2016): “Ministério aponta 1,1 milhão de irregularidades no Bolsa Família,” *G1 Globo*.
- CARDOSO, E. AND A. P. SOUZA (2003): “The impact of cash transfers on child labor and school attendance in Brazil,” *Working Paper*.
- CGU (2014a): “39a Etapa do Programa de Fiscalização a partir de Sorteios Públicos - Baía da Traição/PB,” Tech. rep., Controladoria-Geral da União.
- (2014b): “39a Etapa do Programa de Fiscalização a partir de Sorteios Públicos - Camapuã/MS,” Tech. rep., Controladoria-Geral da União.
- (2014c): “39a Etapa do Programa de Fiscalização a partir de Sorteios Públicos - Xavantina/SC,” Tech. rep., Controladoria-Geral da União.
- (2015): “40a Etapa do Programa de Fiscalização a partir de Sorteios Públicos - Sete Quedas/MS,” Tech. rep., Controladoria-Geral da União.
- CHE, Y.-K. AND I. L. GALE (1998): “Caps on political lobbying,” *The American Economic Review*, 88, 643–651.
- CHEN, B. R. AND Y. S. CHIU (2011): “Competitive bidding with a bid floor,” *International Journal of Economic Theory*, 7, 351–371.
- COHN, A., E. FEHR, AND M. A. MARÉCHAL (2014): “Business culture and dishonesty in the banking industry,” *Nature*, 516, 86.
- COHN, A., M. A. MARÉCHAL, AND T. NOLL (2015): “Bad boys: How criminal identity salience affects rule violation,” *The Review of Economic Studies*, 82, 1289–1308.
- COLONNELLI, E. AND M. PREM (2017): “Corruption and firms: evidence from randomized audits in Brazil,” *SSRN 2931602*.
- COMPTE, O., A. LAMBERT-MOGILIANSKY, AND T. VERDIER (2005): “Corruption and competition in procurement auctions,” *Rand Journal of Economics*, 1–15.
- DAÏEFF, L. (2015): “Why the Bolsa Família is not clientelistic (and what it might be instead),” *Chroniques des Amériques*, 15.
- DANA, J., R. A. WEBER, AND J. X. KUANG (2007): “Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness,” *Economic Theory*, 33, 67–80.
- DE BRAUW, A., D. O. GILLIGAN, J. HODDINOTT, AND S. ROY (2015): “The impact of Bolsa Família on schooling,” *World Development*, 70, 303–316.
- DE JANVRY, A., F. FINAN, AND E. SADOULET (2012): “Local electoral incentives and decentralized program performance,” *Review of Economics and Statistics*, 94, 672–685.
- DI TELLA, R. AND E. SCHARGRODSKY (2003): “The role of wages and auditing during a crackdown on corruption in the city of Buenos Aires,” *The Journal of Law and Economics*, 46, 269–292.
- FAUSTO MACEDO, JULIA AFFONSO, M. C. (2016): “Procuradoria aponta fraude de R\$ 2,5 bi no Bolsa Família,” *Estadão*.
- FERRAZ, C. AND F. FINAN (2008): “Exposing corrupt politicians: the effects of Brazil’s publicly released audits on electoral outcomes,” *The Quarterly Journal of Economics*, 123, 703–745.
- FERRAZ, C., F. FINAN, AND D. B. MOREIRA (2012): “Corrupting learning: Evidence from missing federal education funds in Brazil,” *Journal of Public Economics*, 96, 712–726.

- FIRPO, S., R. PIERI, E. PEDROSO JR, AND A. P. SOUZA (2014): “Evidence of eligibility manipulation for conditional cash transfer programs,” *Economia*, 15, 243–260.
- FISCHBACHER, U. AND F. FÖLLMI-HEUSI (2013): “Lies in disguise – an experimental study on cheating,” *Journal of the European Economic Association*, 11, 525–547.
- FRIED, B. J. (2012): “Distributive politics and conditional cash transfers: the case of Brazil’s Bolsa Família,” *World Development*, 40, 1042–1053.
- GAVIOUS, A., B. MOLDOVANU, AND A. SELA (2002): “Bid costs and endogenous bid caps,” *RAND Journal of Economics*, 709–722.
- GIANNETTI, M., G. LIAO, J. YOU, AND X. YU (2017): “The externalities of corruption: Evidence from entrepreneurial activity in China,” *CEPR Discussion Paper No. DP12345*.
- GLEWWE, P. AND A. L. KASSOUF (2012): “The impact of the Bolsa Escola/Familia conditional cash transfer program on enrollment, dropout rates and grade promotion in Brazil,” *Journal of Development Economics*, 97, 505–517.
- GUPTA, S., H. DAVOODI, AND E. TIONGSON (2001): “Corruption and the provision of health care and education services,” in *The political economy of corruption*, Routledge, 123–153.
- HIDER, J. (2014): “Cat is out of the bad in welfare scam in Brazil,” *The Times*.
- HOLMBERG, S. AND B. ROTHSTEIN (2011): “Dying of corruption,” *Health Economics, Policy and Law*, 6, 529–547.
- HUNTER, W. AND T. J. POWER (2007): “Rewarding Lula: Executive power, social policy, and the Brazilian elections of 2006,” *Latin American Politics and Society*, 49, 1–30.
- KAUFMANN, D. (2005): “Six Questions on the Cost of Corruption with World Bank Institute,” Tech. rep., The World Bank.
- KENT, D. M., E. STEYERBERG, AND D. VAN KLAVEREN (2018): “Personalized evidence based medicine: predictive approaches to heterogeneous treatment effects,” *Bmj*, 363, k4245.
- KRUPKA, E. L. AND R. A. WEBER (2013): “Identifying social norms using coordination games: Why does dictator game sharing vary?” *Journal of the European Economic Association*, 11, 495–524.
- LI, T. AND X. ZHENG (2009): “Entry and competition effects in first-price auctions: theory and evidence from procurement auctions,” *The Review of Economic Studies*, 76, 1397–1429.
- LINDERT, K., A. LINDER, J. HOBBS, AND B. DE LA BRIÈRE (2007): “The nuts and bolts of Brazil’s Bolsa Família Program: implementing conditional cash transfers in a decentralized context,” *Social Protection Discussion Paper*.
- MANI, A., S. MULLAINATHAN, E. SHAFIR, AND J. ZHAO (2013): “Poverty impedes cognitive function,” *science*, 341, 976–980.
- MAURO, P. (1995): “Corruption and growth,” *The quarterly journal of economics*, 110, 681–712.
- (2004): “The persistence of corruption and slow economic growth,” *IMF staff papers*, 51, 1–18.
- MAZAR, N., O. AMIR, AND D. ARIELY (2008): “The dishonesty of honest people: A theory of self-concept maintenance,” *Journal of marketing research*, 45, 633–644.
- MCBRIDE, L. AND A. NICHOLS (2018): “Retooling poverty targeting using out-of-sample validation and machine learning,” *The World Bank Economic Review*, 32, 531–550.

- MCPAKE, B., D. ASHIMWE, F. MWESIGYE, M. OFUMBI, L. ORTENBLAD, P. STREEFLAND, AND A. TURINDE (1999): “Informal economic activities of public health workers in Uganda: implications for quality and accessibility of care,” *Social science & medicine*, 49, 849–865.
- MDS AND SENARC (2015): “Manual de Gestão do Programa Bolsa Família (2a Edição atualizada),” Tech. rep., Governo Federal, Ministério do Desenvolvimento Social e Combate à Fome e Secretaria Nacional de Renda de Cidadania.
- (2018): “Manual de Gestão do Programa Bolsa Família (3a Edição atualizada),” Tech. rep., Governo Federal, Ministério do Desenvolvimento Social e Secretaria Nacional de Renda de Cidadania.
- MEEKER, D., T. K. KNIGHT, M. W. FRIEDBERG, J. A. LINDER, N. J. GOLDSTEIN, C. R. FOX, A. ROTHFELD, G. DIAZ, AND J. N. DOCTOR (2014): “Nudging guideline-concordant antibiotic prescribing: a randomized clinical trial,” *JAMA internal medicine*, 174, 425–431.
- MEEKER, D., J. A. LINDER, C. R. FOX, M. W. FRIEDBERG, S. D. PERSELL, N. J. GOLDSTEIN, T. K. KNIGHT, J. W. HAY, AND J. N. DOCTOR (2016): “Effect of behavioral interventions on inappropriate antibiotic prescribing among primary care practices: a randomized clinical trial,” *Jama*, 315, 562–570.
- MO, P. H. (2001): “Corruption and economic growth,” *Journal of comparative economics*, 29, 66–79.
- MOSTERT, S., F. NJUGUNA, G. OLBARA, S. SINDANO, M. N. SITARESMI, E. SUPRIYADI, AND G. KASPERS (2015): “Corruption in health-care systems and its effect on cancer care in Africa,” *The Lancet Oncology*, 16, e394–e404.
- OGLOBO (2016): “Pente-fino do governo descobre fraude no Bolsa Família,” *O Globo*.
- OLKEN, B. A. (2006): “Corruption and the costs of redistribution: Micro evidence from Indonesia,” *Journal of public economics*, 90, 853–870.
- (2007): “Monitoring corruption: evidence from a field experiment in Indonesia,” *Journal of political Economy*, 115, 200–249.
- OLKEN, B. A. AND P. BARRON (2009): “The simple economics of extortion: evidence from trucking in Aceh,” *Journal of Political Economy*, 117, 417–452.
- PEI, Z., J.-S. PISCHKE, AND H. SCHWANDT (2019): “Poorly measured confounders are more useful on the left than on the right,” *Journal of Business & Economic Statistics*, 37, 205–216.
- PENFOLD-BECERRA, M. (2007): “Clientelism and social funds: Evidence from Chávez’s Misiones,” *Latin American Politics and Society*, 49, 63–84.
- RANDOLPH-SENG, B. AND M. E. NIELSEN (2007): “Honesty: One effect of primed religious representations,” *The International Journal for the Psychology of Religion*, 17, 303–315.
- RAVALLION, M. (2008): “Miss-targeted or Miss-measured?” *Economics Letters*, 100, 9–12.
- (2009): “How relevant is targeting to the success of an antipoverty program?” *The World Bank Research Observer*, 24, 205–231.
- REINIKKA, R. AND J. SVENSSON (2005): “Fighting corruption to improve schooling: Evidence from a newspaper campaign in Uganda,” *Journal of the European economic association*, 3, 259–267.
- SCHAFFLAND, E. (2012): “Conditional Cash Transfers in Brazil: Treatment Evaluation of the Bolsa Família Program on Education,” Tech. rep., Courant Research Centre: Poverty, Equity and Growth-Discussion Papers.

- SHANKAR, S., R. GAIHA, AND R. JHA (2011): “Information, access and targeting: The national rural employment guarantee scheme in India,” *Oxford Development Studies*, 39, 69–95.
- SHU, L. L., N. MAZAR, F. GINO, D. ARIELY, AND M. H. BAZERMAN (2012): “Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end,” *Proceedings of the National Academy of Sciences*, 109, 15197–15200.
- STOEFFLER, Q., B. MILLS, AND C. DEL NINNO (2016): “Reaching the Poor: cash transfer program targeting in Cameroon,” *World Development*, 83, 244–263.
- SUGIYAMA, N. B. AND W. HUNTER (2013): “Whither clientelism? Good governance and Brazil’s Bolsa Família program,” *Comparative Politics*, 46, 43–62.
- TANZI, V. AND H. DAVOODI (1998): “Corruption, public investment, and growth,” in *The welfare state, public investment, and growth*, Springer, 41–60.
- TCU (2006): “Relatório de Acompanhamento do Programa Bolsa Família,” Tech. Rep. TC n° 022.093.2006-5, Governo Federal, Tribunal de Contas da União.
- ZAMBONI, Y. AND S. LITSCHIG (2018): “Audit risk and rent extraction: Evidence from a randomized evaluation in Brazil,” *Journal of Development Economics*, 134, 133–149.
- ZHENG, C. Z. (2001): “High bids and broke winners,” *Journal of Economic theory*, 100, 129–171.
- ZUCCO, C. (2009): “Cash-Transfers and Voting Behavior: An Empirical Assessment of the Political Impacts of the Bolsa Família Program,” in *APSA 2009 Toronto Meeting Paper*.

Chapter 4

Genes, Pubs, and Drinks: Gene-Environment Interplay and Alcohol Licensing Policy in the UK

PIETRO BIROLI AND CHRISTIAN ZÜND

Abstract: Are we genetically destined to behave poorly, or can a well-designed policy and a nurturing environment prevail over our instincts? This paper analyzes the interplay of public policy and individuals' genetic endowments, demonstrating how people's genetic propensity to drink moderates their consumption behavior in response to alcohol availability and licensing policy. We combine data from the UK Biobank with geo-coded data on pubs and retailers, as well as data on alcohol licensing from local authorities in England and Wales. This allows us to construct a fine-grained measure of local alcohol availability for each one of the approximately 500,000 participants in the UK Biobank. Our results show that individuals with a high genetic propensity to drink self-select into environments with easier access to alcohol, react less to changes in the number of sales points, and respond less to restrictive licensing. Importantly, while local licensing authorities are allowed to consider the effects of pubs on children, crime, or public disturbance, they cannot base their decision on public health factors, which mitigates concerns of reverse causality. Using information on physician-diagnosed medical conditions from the National Health Service, we quantify the effect from a public health perspective, showing that the polygenic score predicts alcohol-related disease even when we control for self-reported drinking behavior. Thus, we show that supply-focused licensing policy to mitigate alcohol abuse can clash with individual predispositions and might exacerbate genetic inequality, suggesting the need for a more targeted approach.

Acknowledgments: We are grateful for many helpful discussions including those with Johannes Abeler, Silvia Barcellos, Dan Benjamin, Leandro Carvalho, David Cesarini, Ernst Fehr, Titus Galama, Michel Maréchal, Melinda Mills, Patrick Turley, the GEIGHEI project team, and audiences at the DIAL conference in Turku, University of Oxford, University of Bristol, University of Reading, University of Nürnberg, University of Würzburg, University of Alicante, and University of Zurich. We thank Amr Elriedy, Michelle Rosenberg, and Jeremy Vollen for excellent research assistance. Financial support from NORFACE-DIAL grant 462-16-100 is gratefully acknowledged.

Genes load the gun. Lifestyle pulls the trigger.

— *Dr. Elliott Joslin*

I. INTRODUCTION

Are we genetically destined to behave poorly, or can a well-designed policy and a nurturing environment prevail over our instincts? Debates about the relative influence of nature versus nurture on human behaviors are amongst the oldest in the social sciences (Mulcaster, 1582; Hume, 1748; Darwin, 1859; Freud, 1930). In recent decades, however, it has become increasingly clear that pitting nature against nurture should be relinquished in favor of a more systemic view that considers the complex interplay that may exist between people’s genetic makeup (“nature”) and the environment in which they develop (“nurture”) (Hunter, 2005; Heckman, 2007). In this paper, we evaluate how genetic predisposition and local alcohol licensing policy interact and jointly influence people’s alcohol consumption choices.

Alcohol consumption is worth your attention because of its prevalence and its negative consequences: it is estimated to be the third leading cause of preventable death, it has been related to more than 60 medical conditions, and it accounts for a share of the global burden of disease comparable to those of tobacco or hypertension (Mokdad et al., 2004; Room et al., 2005). In other words, drinking alcohol is one of the leading behavioral factors that contribute to increasing health inequality. But what is the origin of this inequality? Recent research shows that our genes affect how much alcohol we drink, but it is unclear how our genetic predisposition influences our reaction to changes in the availability of alcohol and what this implies for effective alcohol licensing policy.

Our paper shows that the genetic propensity to drink alcohol contributes to health inequalities in two ways: by promoting selection into unfavorable environments, and by decreasing susceptibility to more restrictive licensing policies. Both negative selection and decreased susceptibility lead to higher alcohol intake and, eventually, alcohol-related diseases.

To better understand the genetic origin of differences in health behaviors and inequality, we combine individual genetic data with measures of local alcohol availability for approximately 500,000 participants in the UK Biobank, an extensive and detailed prospective study of British, Welsh, and Scottish participants aged 40 to 69 years recruited between 2006–2010. Using the coordinates of all pubs and the branches of all the major retailers in the UK, we construct a fine-grained measure of local alcohol availability. Using information on 700,000 genetic variants, we estimate a polygenic score that proxies the individual genetic propensity for alcohol consumption.

First, we show that both living in proximity to many alcohol sales points and a high polygenic propensity for alcohol consumption are associated with several drinking-related behaviors. Moreover, we find that individuals with a high polygenic score self-select into environments with greater alcohol availability. This is evidence of considerable gene-environment correlation: carri-

ers of similar genetic variants tend to cluster in the same area. Similar forms of gene-environment correlation have recently been demonstrated to contribute to inequality in health, socioeconomic status, and education (Belsky et al., 2018, 2019; Abdellaoui et al., 2018).

Second, we evaluate the effectiveness of alcohol licensing policy in tackling health inequality. Since the Licensing Act of 2003, the decision of who should be allowed to own a license to sell and distribute alcohol in the United Kingdom is in the hands of 350 local licensing committees. This subsidiarity leads to substantial geographic variation in the restrictiveness of alcohol distribution. Importantly, while the Licensing Act provides local authorities with significant licensing flexibility, it prohibits them from considering public health, much less the genetic predispositions of their residents, when deciding whether to grant a license. This policy limitation mitigates concerns of reverse causality. Using the data on local licensing activity, we estimate the effect of licensing policy on alcohol consumption for individuals with a low and high polygenic score. We find that a more restrictive licensing policy leads to decreased alcohol intake on average, but individuals with a high genetic propensity to drink are less responsive to a policy change. Therefore, this public policy limiting the supply of alcohol tends to amplify existing genetic inequalities: it is more effective for those individuals who already have a low genetic predisposition to drinking, but it has less bite for those individuals who might need it the most.

In the final part, we investigate the implication of our results from a public health perspective, using information on physician-diagnosed medical conditions from records of the National Health Service. Our results show that individuals with a high genetic propensity to drink are significantly more likely to have an alcohol-related condition, including liver diseases, psychological disorders, and various afflictions of the digestive system. These results hold even if we control for self-reported alcohol consumption, demonstrating how incorporating genetic information can help us uncover relevant dimensions of health-inequality that are not commonly observable.

Our results demonstrate how measures of genetic predisposition can shed light on the determinants and the dynamics of health inequalities. Genetic predispositions interact with individual choices of location and health behaviors, and with public health policies. Our results contribute to the recent literature estimating how the genetic endowment of individuals can mute or exacerbate the causal effect of public policies such as the increase of minimum schooling age (Barcellos et al., 2018), mandatory war draft (Schmitz and Conley, 2016b, 2017), or late-career job loss (Schmitz and Conley, 2016a). We show that the effectiveness of supply-focused licensing policy as a tool to mitigate alcohol abuse can clash with individual predispositions and might actually exacerbate genetic inequality, suggesting the need for a more targeted approach.

The paper proceeds as follows: Section II describes our empirical approach. We provide a brief overview of the UK Biobank, our measure of local alcohol availability, and the licensing data. We then describe the construction of the polygenic score for the number of alcoholic drinks per week. Section III presents our results. We show that a higher polygenic score is associated with increased alcohol consumption, self-selection into areas with easier access to alcohol, and

a weaker relationship between the proximity of alcohol sales points and intake. Turning to the effects of licensing policy, we show that individuals with a strong genetic predisposition react less to restrictive licensing. Finally, we show that a high genetic propensity to drink predicts alcohol-related diseases, even if we control for self-reported drinking. Section IV concludes our analysis and discusses its policy implications, limitations, and possible extensions in future research.

II. EMPIRICAL METHODS

This paper combines genetic and alcohol consumption data from the UK Biobank with geocoded data on pubs and retailers that sell alcohol, as well as information on alcohol licensing from all local licensing authorities in England and Wales. Using the coordinates of all pubs and the branches of all the major retailers in the UK, we construct a fine-grained measure of local alcohol availability for each one of the approximately 500,000 participants in the UK Biobank. Our measure of genetic propensity for alcohol consumption is a polygenic score (PGS) for the number of alcoholic drinks per week (DPW).

A. *The UK Biobank*

Genetic information, measures of alcohol intake, and medical data for the participants are collected by the UK Biobank (UKB).¹ Around the time that the first human genome was sequenced in 2003, the UK’s Medical Research Council and the Wellcome Trust launched an ambitious project to collect biological samples from 500,000 people (Collins, 2012). In addition to the genetic information, the UKB also includes data on health-related behaviors, including several measures of alcohol intake, as well as some measures of socioeconomic status (e.g., education, income, occupation). Being tightly integrated with the UK National Health Service (NHS), the UKB also contains extensive information on medical histories and can follow participants long after their initial inclusion. Importantly for the purpose of this paper, health records include data on the location of participants. This combination of genetic information, data on health-related behaviors, and reasonably good geographic information makes the UKB a unique resource to study gene-environment interactions.²

Between 2006 and 2010, approximately 500,000 individuals aged 40 to 60 underwent an initial assessment at one of the UKB’s 23 assessment centers.³ At the initial assessment, par-

1. This paper is part of the Gene-Environment Interplay in the Generation of Health and Education Inequalities (GEIGHEI) collaboration, approved by the UKB as project #41382.

2. Indeed, the ability to studying gene-environment interactions was a key motivation for building a comprehensive biobank and features prominently in the first sentence of the UK Biobank’s protocol: “Scientists have known for many years that our risks of developing different diseases are due to the complex interplay of different factors: our lifestyle and environment; our personal susceptibility (genes); and the play of chance (luck)” (UKB, 2006, p.3).

3. The UKB aims to improve “the prevention, diagnosis and treatment of a wide range of serious and life-threatening illnesses—including cancer, heart diseases, stroke, diabetes, arthritis, osteoporosis, eye disorders,

ticipants completed a series of questionnaires and a computer-assisted interview. Professional healthcare practitioners collected physical and functional measures as well as blood, urine, and saliva samples. Over the last ten years, subsets of participants completed additional assessments: a follow-up assessment (2012 to 2013), an imaging assessment with magnetic resonance imaging of the brain, heart, and body, ultrasound of the carotid arteries, and X-ray absorptiometry imaging of bones and joints (since 2014), repeated 24hr dietary recalls (2011 to 2012), and a mental health module (2016 to 2017).

While the UKB was conceived as a prospective study with an initial cross-section of measurements and subsequent observation of medical histories (Sudlow et al., 2015), the additional assessments enable us to observe alcohol intake at multiple points in time for a subset of participants. Table A4.1 in Appendix A4.II displays the summary statistics, including the number of observations, for drinking-related behaviors in each of the assessments. Reassuringly, these variables are highly correlated (see Table A4.2 in Appendix A4.II.)

Concerning the UKB’s geographic resolution, there are two noteworthy caveats. First, the location of assessment centers was chosen such that at least 150,000 potential participants in this age group live within less than 10 miles of each center (UKB, 2006, p.49). As a result, the vast majority of participants live in urban areas (see Figure A4.1 in Appendix A4.I for a map of UKB participants). Second, information on a participant’s location is rounded to a 1000m grid for privacy reasons. While this is considerably more precise than similar data sources that only provide information on the state or municipality of residence, it has important implications for the construction of our individual-level measure of alcohol availability.

B. Alcohol Availability

Our primary measure for alcohol availability is the number of pubs located within 1000m of a participant’s residence. This measure of pub density was constructed using a comprehensive list of pubs and their location from The Good Pub Guide. The the UK’s longest-running publication of this kind, The Good Pub Guide has been printed since 1982 and is arguably the most complete list of all pubs in the UK.⁴ In addition to reviews from its staff, The Good Pub Guide enables license holders to provide additional information on their businesses and allows registered and verified users to suggest other pubs for review and inclusion. As a result, it is the most comprehensive and up to date list, containing information on the locations of 52,500 pubs and similar venues in the UK.⁵

depression and forms of dementia” (UK Biobank, <https://www.ukbiobank.ac.uk>, accessed on December 18, 2019). To this end, the UK Biobank focuses on healthy individuals in age groups that are at a high risk of developing one of these conditions throughout the study.

4. It also maintains an extensive online presence since 2009 <https://thegoodpubguide.co.uk>, accessed on October 18, 2018.

5. The British Beer and Pub Association counts 48,350 pubs. The Campaign for Real Ale (CAMRA) uses data on 47,500 pubs in their reports. Using a more restrictive definition of pubs, the Office of National Statistics estimates that there are only around 39,000 pubs in the UK (ONS, 2018). As our measure aims to quantify the

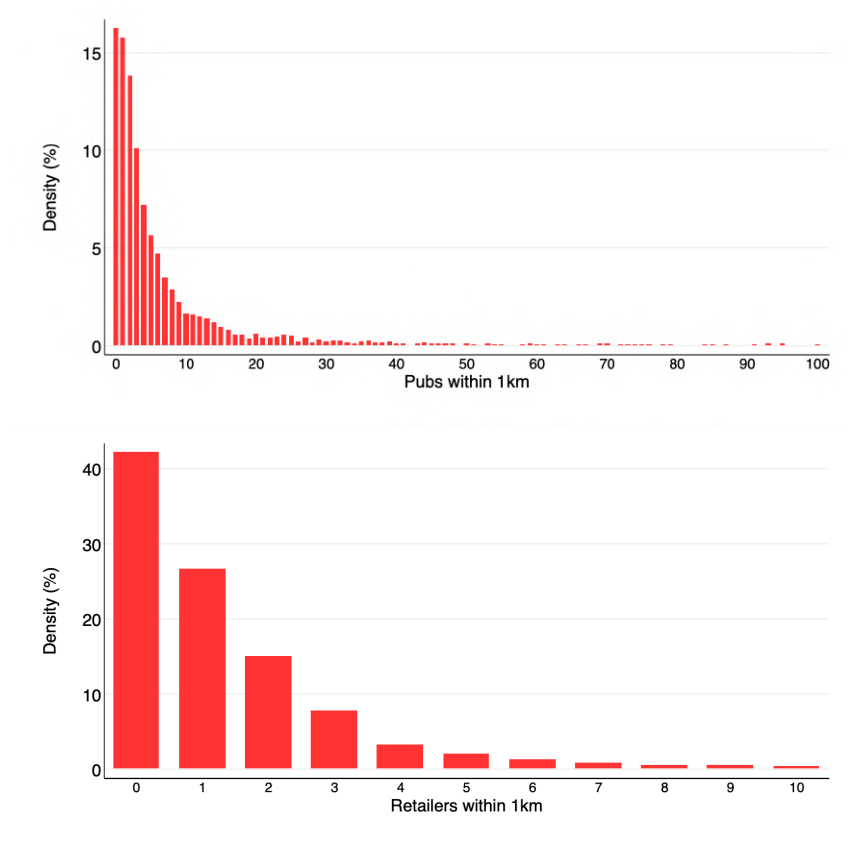


FIGURE 4.1

Distribution of the Alcohol Availability Measures

Notes. This figure displays the distribution of our alcohol availability micro measures. The top panel shows the distribution for the number of pubs within 1000 m of a UKB participant's place of residence (truncated at 100). The bottom panel shows the distribution for the number of retailers (truncated at 10).

While pubs have a prominent place in the UK's drinking and leisure culture, buying alcohol outside of a pub might be cheaper. The on-trade price for a pint in a pub has increased much more quickly than the off-trade price for beer in supermarkets, which has contributed to the growing share of alcohol purchased outside of pubs, culminating in 2015, when off-trade sales overtook on-trade sales for the first time (BBPA, 2016). Therefore, we construct a similar measure for the density of retailers within 1000m, using a panel of the locations of the UK's largest retailers since 2014.⁶

As shown in Figure 4.1, 83.9% of participants live within 1km of at least one pub, and 58.2% live within 1km of a major retailer. These relatively high numbers are not surprising given the concentration of UKB participants in urban areas. There is significant variation in

accessibility of alcohol, a broader definition is more appropriate for our analysis.

6. These include all locations of Tesco, Sainsbury's, Co-op Food, M&S, ASDA, ALDI, LIDL, Morrisons, Waitrose, and a number of regional chains. The data is provided by Geolytix, <https://geolytix.co.uk>, accessed on November 22, 2018.

pub density both between and within local authorities. The average participant living in the City of London⁷ gets to choose between more than 450 venues within 1000m. In contrast, none of the 189 participants in the Richmondshire local authority has more than three pubs in their neighborhood, averaging just 0.61 pubs within a 1000m radius. In addition to the differences between local licensing areas, there is also substantial heterogeneity within areas. For example, in the area governed by the Westminster City Council, pub density ranges from 6 to 408 pubs within a 1000m radius. In general, differences between local authorities explain just under 30% of the variance in pub density.

C. Local Licensing Policy

Since the Licensing Act of 2003, responsibility for licensing in England and Wales has been decentralized to 350 local licensing authorities.⁸ This has led to considerable geographic variation in the restrictiveness or permissiveness of policy.

Each local licensing authority appoints a committee of 10 to 15 members who determine licensing policy in consultation with other stakeholders such as the police or fire and rescue authorities. The licensing committee outlines its strategy for a five year period and conducts regular evaluations. While regulating the sale of alcoholic beverages is a central part of the Licensing Act, the responsibility of licensing committees extends to “licensable activities” more broadly.⁹ In addition to granting, reviewing, refusing, or revoking licenses, the licensing committees have several other tools at their disposal. They can restrict the sale of alcohol to purchases for consumption on-premise or off-premise only, they can limit opening and sales hours, and they can grant the license for entertainment only but prohibit the sale of alcohol entirely. They can also define cumulative impact areas where stricter licensing rules apply, for example, in the vicinity of schools. Applicants for a license to sell alcohol within one of these zones need to prove that they will not negatively impact the objectives of the licensing policy (Martineau et al., 2013).

While the Licensing Act provides local authorities with significant flexibility to shape their policy, it puts strict limits on the factors they can and cannot take into account in their decision making. The act requires licensing committees to consider four specific objectives: the prevention of crime and disorder, public safety, the prevention of public nuisance, and the protection of

7. London’s historic center, governed by the City of London Cooperation and local authority. However, the four local authorities where participants have the next highest average pub density are all also located in central London: the City of Westminster, Camden, Islington, and Kensington and Chelsea.

8. The area of responsibility for a licensing committee usually corresponds to the area of a local government—non-metropolitan districts, metropolitan districts, London’s boroughs, or unitary authorities responsible for the provision of local services. For convenience, we will usually refer to these areas as local authorities. Alcohol licensing law differs in Scotland, where they are governed by the Licensing (Scotland) Act 2005, and in Northern Ireland, which has a fixed quota for the number of pubs and new businesses have to wait for an existing premise to surrender its license.

9. These include regulated entertainment such as performances of plays, dances, live and recorded music, boxing matches, and indoor sports and entertainment more generally.

children from harm. Importantly for the purpose of our analysis, local licensing authorities in England and Wales are not allowed to take public health into consideration, which mitigates concerns of reverse causality.¹⁰ This is made explicit by the Home Office in its binding guidance for local authorities:

“Any conditions imposed must be appropriate for the promotion of the licensing objectives; there is no power for the licensing authority to attach a condition that is merely aspirational. For example, conditions may not be attached which relate solely to the health of customers rather than their direct physical safety” (Home Office, 2012, p.74).¹¹

Although restricting licensing committees to these four objectives does not imply that licensing policy is exogenous, it does at least imply that policy is not explicitly conditioned on the (expected) genetic predisposition to consume alcohol. Figure A4.4 in Appendix A4.I suggests that this is indeed not the case; policy measures are uncorrelated with our measure of genetic predisposition of UKB participants living in the local authority. Moreover, Table A4.3 in Appendix A4.II shows that these policy measures are well-balanced between local authorities with an average genetic propensity below and above the median. As a more stringent test of balancedness, we conducted a Pei et al. (2019) test to check whether licensing policy is predictive of the local genetic predisposition. This is not the case ($\chi^2(3) = 3.500$, $P = 0.321$).¹²

D. Polygenic Scores (PGS)

Following the most recent literature in social science genomics (Benjamin et al., Forthcoming; Braudt, 2018; Freese, 2018), individuals’ genetic predisposition for alcohol consumption is quantified using a polygenic score (PGS), a summary of the effects of more than 700,000 genetic variants that contribute to an individual’s tendency to drink. The polygenic score employed in this paper predicts drinking based on a weighted sum of single-nucleotide polymorphisms (SNPs), the most common form of genetic variation present in the human genome.¹³

10. This is not the case in Scotland. In addition to the four objectives in England’s licensing act, Section 4 of the Licensing (Scotland) Act 2005 lists “protecting and improving public health” as an additional objective (Scottish Parliament, 2005). This public health provision increases the scope for endogeneity, as licensing policy might be determined in response to local alcohol consumption and potentially to counteract a suspected genetic predisposition. We thus focus on England and Wales in our analysis.

11. Naturally, licensing conditions imposed for reasons other than public health might nevertheless have a positive impact on public health. However, this should not be the objective of the policy: “There will of course be occasions when a public safety condition could incidentally benefit a person’s health more generally, but it should not be the purpose of the condition as this would be outside the licensing authority’s powers (be ultra vires) under the 2003 Act.” (Home Office, 2012, p.11)

12. The test estimates the coefficients when the policy variable is regressed on the polygenic score, the first 40 principal components of the genetic data, and the fixed effects: $X_{i,s} = \alpha + \beta PGS_i + PC_i + f(age_i, sex_i) + \varepsilon_{i,s}$. It then uses an F-test whether these coefficients are jointly significant from zero. A significant test statistic would imply that these policy variables together can predict the local genetic predisposition.

13. The human genome is a series of about 3 billion letter pairs (A,G,T,C) which compose an individual’s DNA code. While most of the DNA sequence is the same from one person to the next, there are some parts

The polygenic score is constructed in the following way:

$$PGS_i = \sum_{j \in J} \hat{\beta}_j^{GWAS} g_{i,j},$$

where $g_{i,j} \in \{0, 1, 2\}$ is individual i 's genotype (the number of reference alleles) at SNP j , and $\hat{\beta}_j^{GWAS}$ is the weight of SNP j estimated from a genome-wide association study (GWAS) of the form:

$$Y_i = \alpha + \beta_j^{GWAS} g_{i,j} + f(\text{age}_i, \text{sex}_i) + PC_i + \varepsilon_{i,j}$$

where the outcome of interest Y_i is regressed on a single SNP j , (non-parametric) controls for age and sex, and the first 40 principal components of the genetic data. This regression is run J times, one per each SNP. We use $\hat{\beta}_j^{GWAS}$ as weights in our polygenic score.¹⁴ The polygenic score is normalized to have mean zero and standard deviation one to facilitate interpretation.

As there are no sufficiently powered publicly available GWAS summary statistics for alcohol consumption that do not include the UKB, we run our own GWAS to construct the weights. To mitigate the problem of overfitting that arises whenever individuals are part of both the sample used to calculate the weights and the sample for polygenic prediction, we use a cross-fitting approach. We create ten random folds of the sample \mathcal{F}_k , $k \in \{1, \dots, 10\}$.¹⁵ For each fold \mathcal{F}_k , we construct a polygenic score using weights $\hat{w}_{-k,j}$ obtained from running a GWAS on the other 90% of the sample \mathcal{F}_{-k} :

$$PGS_i = \sum_{j \in J} \hat{w}_{-k,j} g_{i,j}, \quad \forall i \in \mathcal{F}_k$$

Our polygenic score predicts *Drinks per Week*, defined by Liu et al. (2019) as the average number of drinks a participant reported drinking, summed over all types of alcohol. The outcome is log-transformed using $\log(x + 1)$ prior to the GWAS to reduce the effect of outliers on the estimated weights. Figure 4.2 displays the distribution of the polygenic score for UKB participants and a binned scatter plot that illustrates how the average number of drinks per week increases for individuals with a higher polygenic score.

of the DNA that differ across people. The most common form of genetic variation are SNPs, Single Nucleotide Polymorphisms, and they represent a difference in a single nucleotide (one DNA-“letter”). There are roughly 10 million SNPs in the human genome. Each SNPs has two alleles, i.e. two letters, one for each chromosome: the one inherited from the dad, and the one inherited from the mom. These two letters identify your genotype, the genetic type of a particular person at a precise genetic locus.

14. Alternatively, the estimated coefficients $\hat{\beta}_j$ could be adjusted for linkage disequilibrium—the correlation between nearby SNPs—using LDpred (Vilhjálmsdóttir et al., 2015) to obtain the weights w_j .

15. This approach is discussed in (Mak et al., 2018). A similar approach is used in (Machiela et al., 2011). See Dudbridge (2013) for a critical review of this and other methods to increase the power of polygenic scores.

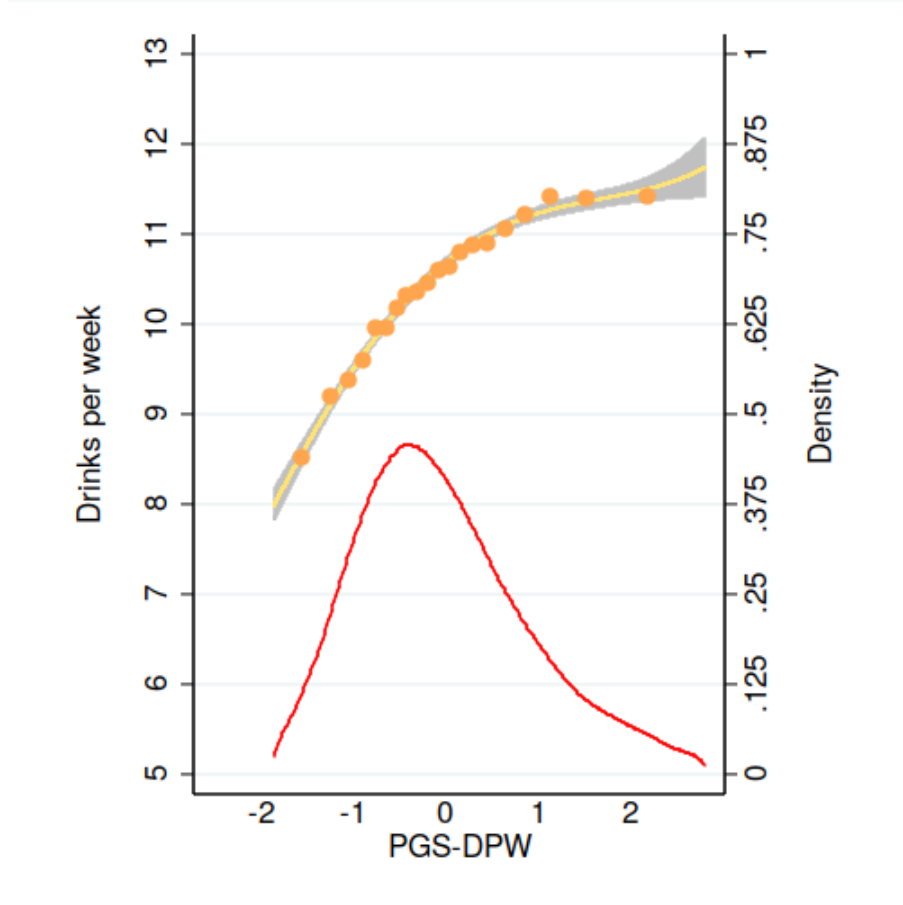


FIGURE 4.2

Distribution and Predictive Power of the Polygenic Score

Notes. This figure displays the distribution of the polygenic score and the expected number of drinks per week for UKB participants with different levels of genetic predisposition. The density function is truncated at the 1st and 99th percentile of the polygenic score. The orange points represent the mean number of drinks per week and the mean polygenic score for 20 equally sized bins of UKB participants. The yellow line and the shaded gray area show a fitted polynomial of degree three and the associated 95% confidence interval.

E. Empirical Specification

We run three main analyses: the first to validate that the PGS predicts drinking (prediction); the second to evaluate the extent of self-selection into unhealthy locations (gene-environment correlation); the third to understand whether the PGS moderates the effect of alcohol availability (gene-environment interaction).

First, we validate that individuals with a strong genetic predisposition indeed drink more. We run the following regression:

$$Y_{i,s} = \alpha + \beta \text{PGS}_i + PC'_i \theta + f(\text{age}_i, \text{sex}_i) + \nu_s + \varepsilon_{i,s} \quad (4.1)$$

where $Y_{i,s}$ are different measures of alcohol intake for UKB participant i living in local authority s , and PGS_i is the polygenic score for individual i —which is constant over time regardless of where they live.

To test the robustness of our results, we control for genetic stratification, demographic differences, and local authority fixed effects. We use the first 40 principal components PC_i of the full genetic data to account for systematic differences in ancestry and population stratification, i.e., the difference in allele frequencies between (sub)populations which might be spuriously associated with the outcome of interest (Price et al., 2006; Rietveld et al., 2014). To rule out that our findings simply reflect well-documented age and gender patterns in alcohol consumption (e.g., Wilsnack et al., 2000; Grant et al., 2004; Wilsnack et al., 2009), we non-parametrically control for $f(age_i, sex_i) = Age \times Sex$ indicators. Our most comprehensive model also includes local authority fixed effects ν_s to account for the considerable heterogeneity in alcohol availability and other unobservable regional differences.¹⁶ The error term $\varepsilon_{i,s}$ is allowed to cluster at the local authority level, where the alcohol licensing policy is determined.

Second, we address the question of whether participants with a high genetic propensity to drink self-select into environments with more alcohol sales points, a measure of gene-environment correlation. We run the following regression:

$$E_{i,s} = \alpha + \beta PGS_i + PC_i' \theta + f(age_i, sex_i) + \nu_s + \varepsilon_{i,s} \quad (4.2)$$

where $E_{i,s}$ are different measures of alcohol availability in i 's environment in local authority s , such as the number of pubs, retailers, or alcohol sale licenses. Again, depending on the specification, we add different levels of controls.

Third, in our principal analysis, we test to which extent the polygenic score moderates the association between the environment—the density of alcohol sales points or various licensing policies—and alcohol consumption. This allows us to estimate the extent of gene-environment interaction. We do this by regressing the outcome of interest on the polygenic score, a measure of the environment, and the interaction of the two using the following econometric specification:

$$Y_{i,s} = \alpha + \beta PGS_i + \gamma E_{i,s} + \delta (PGS \times E)_{i,s} + PC_i' \theta + f(age_i, sex_i) + \nu_s + \varepsilon_{i,s} \quad (4.3)$$

where the $Y_{i,s}$, the outcome of interest for UKB participant i living in local authority s , is regressed on the polygenic score PGS_i , a measure of the environment $E_{i,s}$, and their interaction $(E \times PGS)_{i,s}$. The coefficient δ measures the extent to which the polygenic score moderates the association between the environment and the outcome of interest. Again, depending on the specification, we add different levels of controls.

16. We cannot include local authority fixed effects in our analysis of policy effectiveness, because licensing policy is set at the local authority level.

III. RESULTS

A. *Genetic Predisposition and Alcohol Consumption*

Figure 4.2 already shows that individuals with a higher PGS report drinking more alcoholic beverages; this is reassuring but not novel: it merely shows that the score can predict the outcome it was constructed to predict. Continuing from this, we test whether the polygenic score is predictive of the other measures of alcohol consumption in the UKB.

Figure 4.3 shows the effects of a one standard deviation increase in the PGS on each of the alcohol-related outcomes in the UKB, using false-coverage rate adjusted confidence bands (Benjamini and Hochberg, 1995) to account for multiple hypothesis testing. The PGS is significantly associated with each outcome at the 99% level.¹⁷ Several things are noteworthy:

First, the results are very much as expected. Participants with a higher polygenic score drink more, drink more frequently, and are more likely to go to the pub at least weakly. Even among participants who don't drink alcohol, those with a higher PGS are significantly more likely to report drinking in the past. Questions from the mental health module show that respondents with a high PGS not only drink more often but also drink more in a typical drinking spell, and that they report more problems with addictive behavior. The only outcome that is negatively associated with the PGS measures whether participants usually only drink alcohol with a meal, where, unsurprisingly, scoring high on the PGS is associated with more drinking outside of mealtimes.

Second, the associations are consistent and robust. For most outcomes, one standard deviation increase in the PGS predicts a 0.1-0.2 standard deviation change in the outcome variable. The results are consistent, no matter whether participants complete a comprehensive 24h dietary recall or whether they simply report about an average week. They hold for traits measured on very short time-horizons (e.g., "Drank yesterday") and over much more extended periods (e.g., "Drinks more than 10 years ago").

Third, the effects are stable in later rounds of the assessment, even-though the PGS has been constructed from drinking behavior around at the time of the first assessment. This stability is particularly important as a person's genetic endowment is fixed at conception. While there might be some variation in how specific genes influence an individual's behavior over the life-cycle, we would not expect much change for the UKB cohort, aged 40 to 69 at the first assessment. Indeed, the fact that the polygenic score is equally able to predict alcohol consumption later in life suggests that it is indeed capturing a stable genetic component of these traits.

17. Figure A4.3 in Appendix A4.I displays the increase in R^2 from including the PGS on top of non-parametric controls for age and gender.

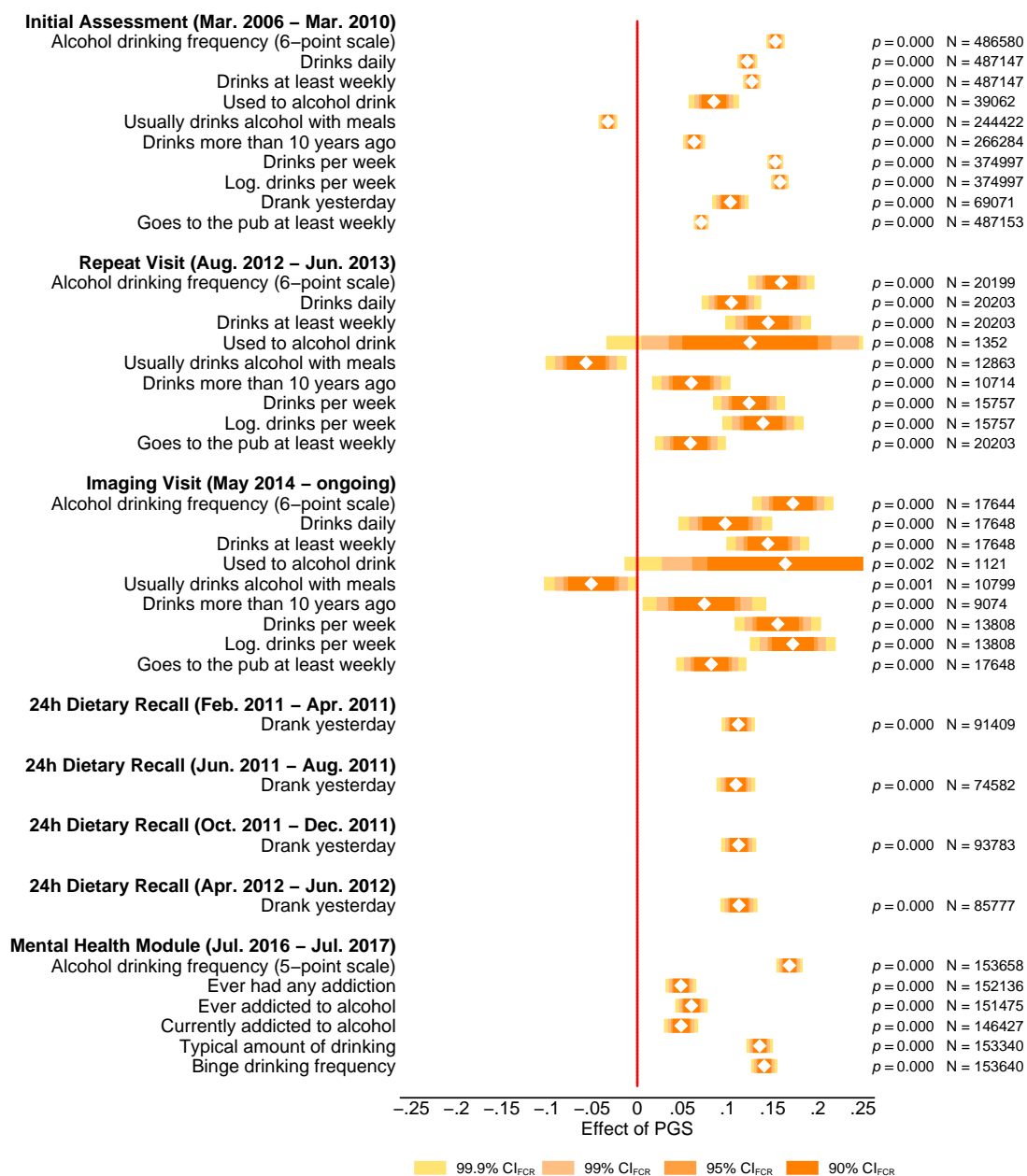


FIGURE 4.3

Relationship between the Polygenic Score and Drinking Behavior

Notes. This figure displays the relationship between the polygenic score and drinking-related outcomes in the UKB. The white diamonds show the estimated effect from 38 regressions where drinking-related outcomes in the UKB are regressed on the polygenic score for the number of drinks per week, the first 40 principal components of the genetic data, “Age × Sex” fixed effects, and local authority fixed effects. Standard errors are clustered at the local authority level. Colors indicate false coverage rate adjusted 90%, 95%, 99%, and 99.9% confidence intervals (Benjamini and Yekutieli, 2005).

B. Genetic Predisposition and Self-Selection

Any analysis of the local environment on individuals' behaviors needs to acknowledge that many people self-selected their place of residence from the options that were available to them. Consider a genetic variant that is associated with higher educational attainment: as universities and career opportunities for university graduates tend cluster in metropolitan areas, we would expect the variant to be more common in urban samples. These forms of gene-environment correlation (Plomin et al., 1977) have been shown for genetic markers associated with education (e.g., Belsky et al., 2019), BMI (e.g., Haworth et al., 2018), substance use (e.g., Harden et al., 2008) and other behaviors (e.g., Abdellaoui et al., 2018).

The top left panel of Figure 4.4 shows that there is indeed a correlation between the number of pubs within 1000m of participants and their PGS. On average, a participant with a PGS one standard deviation below the mean lives around 7.5 pubs. In contrast, participants with a PGS one standard deviation above the mean average 9.0 pubs in their neighborhood. To test whether UKB participants actively select into areas with high pub density, we first compare the number of pubs at their current place of residence with the number of pubs around their place of birth. The top right panel of Figure 4.4 illustrates that while the PGS varies relatively little with the pub density at the birthplace, there is a strong gradient with respect to the pub density at their current place of residence. The results for the number of retailers are comparable. On average, a participant with a PGS one standard deviation below the mean lives around 1.25 retailers, whereas a participant with a PGS one standard deviation above the mean has about 1.5 (Figure 4.4, bottom left). Although the smaller number of major retailers leads to a somewhat patchy heat map (Figure 4.4, bottom right), we can again see that the PGS correlates more with the retailer density at the current place of residence.

Table 4.1 shows that the ratio between the number of alcohol sales point at their place of residence and their place of birth is significantly higher for individuals with a high PGS. At the time of the UKB assessment, a participant with a polygenic score one standard deviation above the mean lives around 11.6% more pubs relative to his place of birth ($P = 0.018$). When we control for the first 40 principal components of genetic stratification and for demographic characteristics, the ratio shrinks to 8.2%, but the effect continues to be highly significant ($P = 0.006$). The same is true for the number of retailers, where a polygenic score one standard deviation above the mean is associated with 3.2% more retailers relative to the place of birth without controls ($P = 0.001$), and 1.7% more retailers with controls ($P = 0.002$). This suggests that participants with a high PGS actively select into areas with easier access to alcohol.

As a final test, we investigate whether a mismatch between a participant's PGS and the pub density around their place of residence at the time of the first UKB assessment (i.e., someone with a high PGS living in a neighborhood with very few pubs or vice versa) predicts whether they have since moved to a new location. Table A4.5 in Appendix A4.II shows that there is

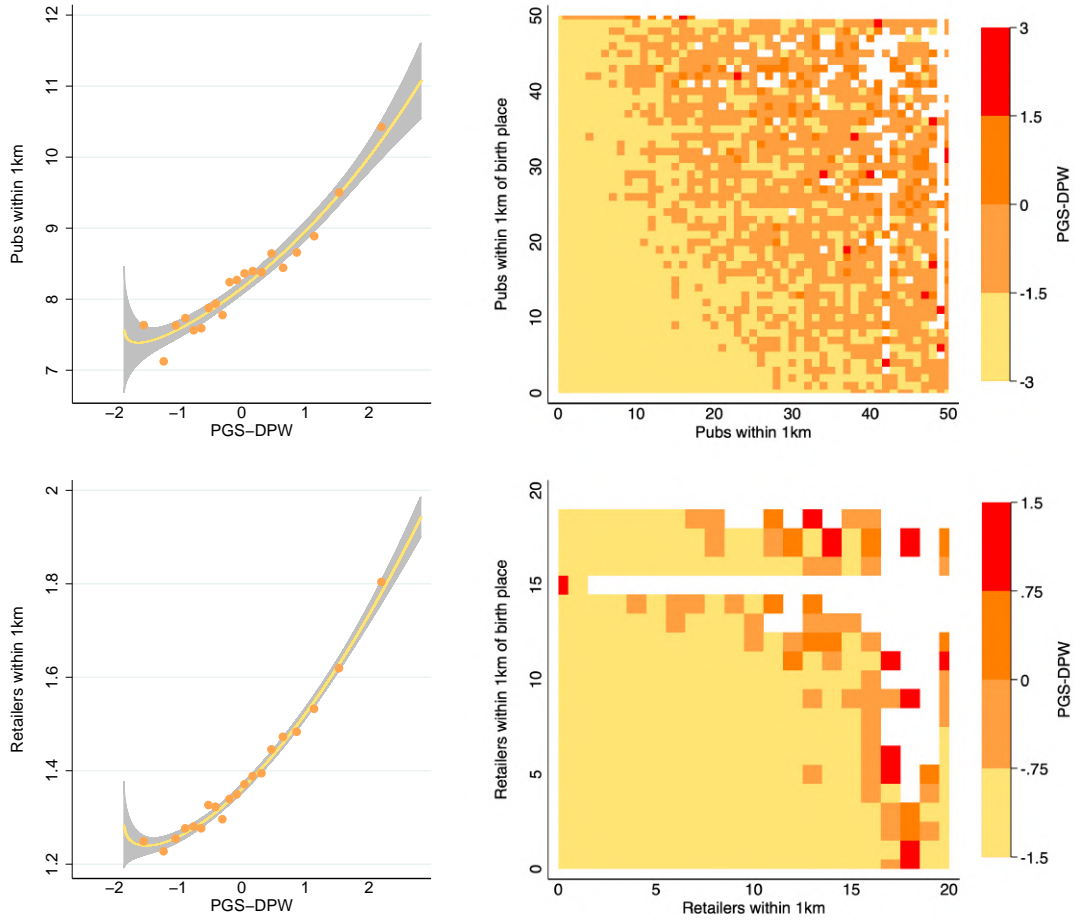


FIGURE 4.4
Evidence of Gene-Environment Correlation

Notes. This figure displays the relationship between the polygenic score and the self-selection of UKB participants into areas with high and low alcohol availability. The top row shows results for the density of pubs within 1000m of UKB participants. The bottom row shows results for the density of retailers. Panels on the left illustrate the relationship between the number of alcohol sales points and participants' genetic propensity. The orange points represent the mean number of sales points and the mean polygenic score for 20 equally sized bins of UKB participants. The yellow line and the shaded gray area show a fitted polynomial of degree three and the associated 95% confidence interval. Panels on the right show a heat map of the genetic propensity for different combinations of the number of sales points around the current location and the number of sales points around the place of birth. Lighter colors indicate lower values of the polygenic score; darker colors indicate stronger predisposition.

indeed a small interaction effect that is consistent with the hypothesis that gene-environment mismatch predicts moving. The interaction is even stronger for retailers (see Table A4.6 in Appendix A4.II).

The finding that individuals seem to select their place of residence partially for proximity to pubs might initially seem unlikely. The willingness to move for access to public goods such as schools and higher education, or to relocate for employment is well-documented (e.g., Nechyba, 2000; Banzhaf and Walsh, 2008; Hensen et al., 2009) and unsurprising given the potentially

TABLE 4.1
SELF-SELECTION INTO AREAS WITH HIGHER ALCOHOL AVAILABILITY

	Ratio of pubs			Ratio of retailers		
	(1)	(2)	(3)	(4)	(5)	(6)
PGS-DPW	0.115* (0.049)	0.081** (0.030)	0.082** (0.030)	0.032** (0.010)	0.017** (0.005)	0.017** (0.005)
Constant	1.436*** (0.131)	1.443*** (0.130)	1.435*** (0.128)	0.575*** (0.035)	0.585*** (0.034)	0.581*** (0.034)
First 40 PCs	No	Yes	Yes	No	Yes	Yes
Age X Sex	No	No	Yes	No	No	Yes
R2	0.000	0.001	0.002	0.001	0.005	0.006
N	371687	371687	371686	318381	318381	318380

Notes. This table reports the effect of genetic predisposition on the ratio of alcohol availability at a participant's place of residence and place of birth. Columns (1) to (3) report the effect on the ratio between the number of pubs within 1000m of a participant's place of residence and their place of birth, Columns (4) to (6) on the ratio between the number of retailers. "PGS-DPW" denotes the polygenic score for the number of drinks per week. Columns (1) and (4) include no additional controls. Columns (2) and (5) control only for genetic stratification using the first 40 principal components of the genetic data. Columns (3) and (6) control for genetic stratification and additionally include "Age \times Sex" fixed effects. Standard errors are clustered at the local authority level. Significance levels: $^+P < 0.1$, $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

considerable gains in life-time earnings (Bowles, 1970; Yankow, 2003; Kidd et al., 2017). But surely few students would turn down a place at the University of Oxford (pub density¹⁸: 15.2; THE Ranking¹⁹: 1st) solely for the higher pub density around the University of Brighton (pub density: 41.0; THE Ranking 600-800th)? However, given the local nature of our measure—pubs within 1000m—we would expect that self-selection happens at the level of intra-urban mobility (Simmons, 1968; Brown and Moore, 1970). Thus, a better way to think about this decision is whether a student with a place at the University of Oxford would rather live in the lively Jericho neighborhood (pub density: 87) or in quiet, residential Marston (pub density: 1).

C. Genetic Predisposition and Elasticity of Demand

Figure 4.5 displays the relationship between the number of pubs within 1000m and normalized alcohol consumption for participants with a PGS above and below the median. We consider four drinking-related outcomes: the self-reported drinking frequency on a 6-point scale (top left), whether a participant reports drinking each day (bottom left), the number of alcoholic drinks per week (top right), and the amount drunk in a typical session (bottom right).

UKB participants who live in an area with a high number of pubs drink significantly more than participants in areas with a low pub density. The elasticity of consumption with respect to availability, however, is considerably lower for participants with a high genetic propensity to drink. In areas with almost no pubs, the difference between the average consumption of the

18. The average number of pubs around UKB participants in the local authority.

19. Times Higher Education Ranking 2020.

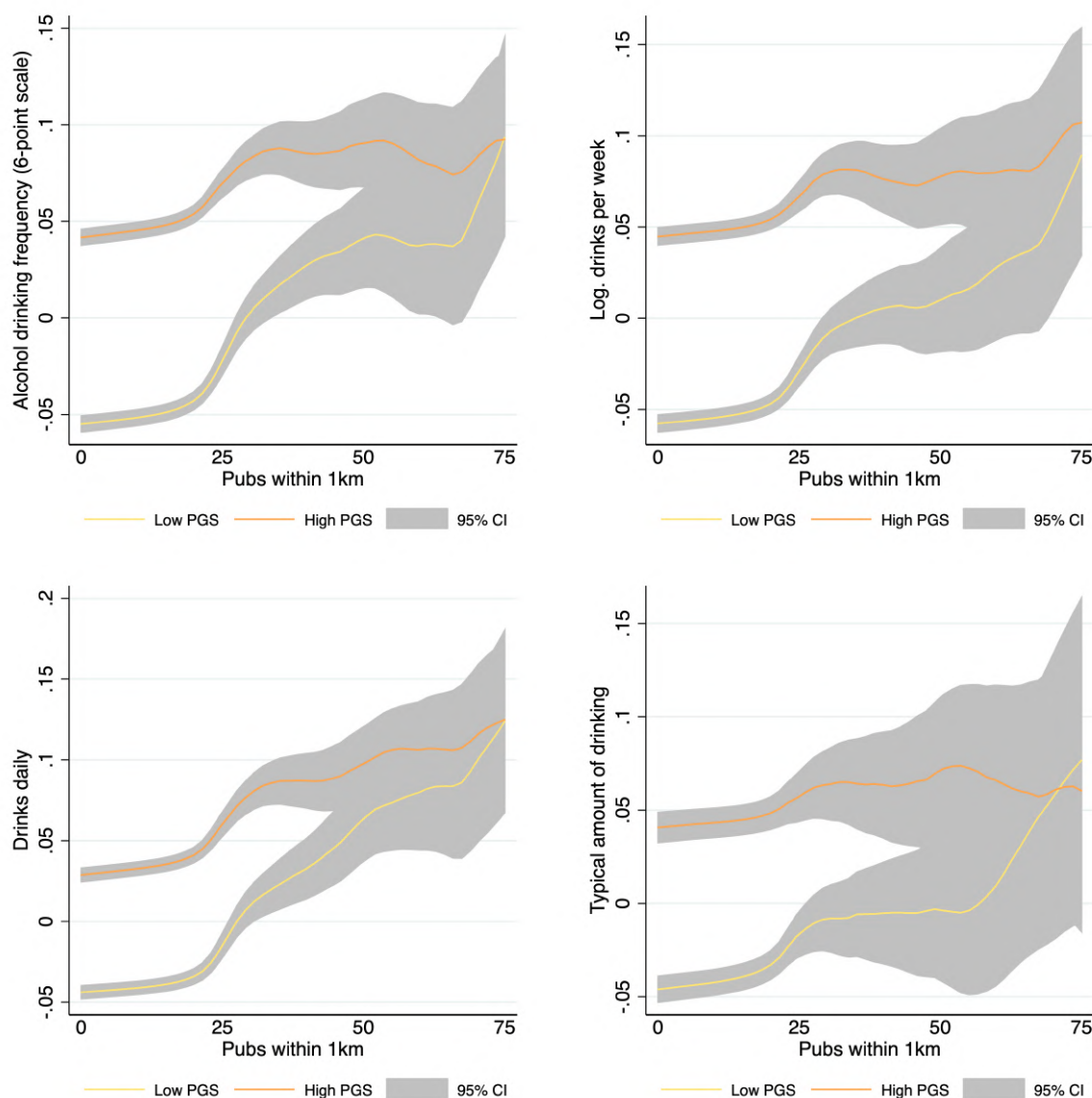


FIGURE 4.5
Drinking Behavior and Alcohol Availability by Polygenic Score

Notes. This figure displays the relationship between drinking behavior and the number of pubs within 1000m of UKB participants with a polygenic score above and below the median. The orange and yellow lines and the shaded gray areas show the polynomial fit and the associated 95% confidence interval for UKB participants with high and low polygenic scores, respectively. The figure was constructed using kernel-weighted local polynomial smoothing. All models control for genetic stratification using the first 40 principal components of the genetic data, “Age \times Sex” fixed effects, and local authority fixed effects.

two groups is roughly one standard deviation for the four outcomes. This gap shrinks as we consider areas with higher and higher pub density; in areas with 75 pubs within 1000m, the average consumption of the two groups is essentially the same. Table A4.4 in Appendix A4.II shows that both the number of pubs and the PGS are significantly positively correlated with

alcohol intake, but that there is a negative and significant interaction effect.

Figure A4.5 in Appendix A4.I shows that the pattern is similar if we instead use the number of retailers within 1000m. As with pub density, UKB participants with a low genetic predisposition who live in an area with a large number of retailers have significantly higher alcohol intake than similar individuals in an area with fewer retailers. For retailers, consumption seems to decrease slightly as the number of sales points increases from zero to one, before increasing again as we consider areas with more and more retailers. Table A4.4 again confirms the positive correlation of drinking with retailer density and the PGS, as well as the existence of a significant negative interaction.

Although these estimates are likely affected by the reported self-selection, the lower elasticity for high PGS participants is in line with the documented low price-elasticity for alcoholic beverages due to their addictive nature (Anderson and Baumberg, 2006).

D. Genetic Predisposition and Licensing Policy

Having observed that individuals with a high genetic propensity for alcohol consumption drink more, self-select into environments with greater alcohol availability, and have less elastic demand for alcohol consumption, we now turn to the question of what these behaviors imply for effective alcohol licensing policy. If individuals with a high genetic predisposition react much less to changes in the availability of alcohol, supply-focused alcohol policy—reducing the number of licenses, restricting hours of sales, or even taxation—might reduce alcohol consumption on average, but have only a negligible effect on the individuals who would benefit most from drinking less.

Tables 4.2, 4.3, and 4.4 show the effect of the number of premise licenses, no-sales licenses, and 24h supermarket licenses on the same four (normalized) outcomes: the self-reported drinking frequency, whether a participant reports drinking each day, the number of alcoholic drinks per week, and the amount drunk in a session. In each case, we first test the effect of the policy on its own, before including the polygenic score and its interaction with the policy measure. For each outcome and licensing measure, the interaction term counteracts the policy—often significantly so. Thus, we find that a higher PGS is associated with a smaller reaction to the policy.

We first test whether the number of premises with a license to sell alcohol affects the drinking behavior of UKB participants in the local authority (Table 4.2). As local authorities differ in population, we use the number of licensed per 1,000 residents as our policy measure. A more lenient licensing policy, characterized by more licensed premises, is associated with increased alcohol consumption. Self-reported drinking frequency increases significantly ($P = 0.022$), as does the proportion of participants that report drinking daily ($P = 0.015$), and the number of drinks per week ($P = 0.035$), but not the typical amount ($P = 0.404$). Turning to the interaction models, we find that, unsurprisingly, the PGS is significantly associated with increased alcohol consumption for all outcomes. More interestingly, UKB participants with a

TABLE 4.2
GENE-ENVIRONMENT INTERACTION FOR LICENSING POLICY (LICENSED PREMISES)

	Drinking frequency		Daily drinker		Log. drinks per week		Typical amount	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
On-premise licenses (per 1,000)	0.237* (0.106)	0.241* (0.104)	0.268* (0.114)	0.271* (0.111)	0.177* (0.085)	0.177* (0.084)	0.016 (0.029)	0.022 (0.026)
PGS-DPW		0.168*** (0.004)		0.135*** (0.004)		0.171*** (0.004)		0.143*** (0.005)
On-premise licenses (per 1,000) \times PGS-DPW		-0.060** (0.019)		-0.049** (0.016)		-0.037** (0.011)		-0.032*** (0.007)
Constant	-0.049* (0.022)	-0.048* (0.021)	-0.044+ (0.022)	-0.043* (0.022)	-0.045** (0.017)	-0.047** (0.017)	-0.013+ (0.008)	-0.014+ (0.007)
First 40 PCs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age \times Sex	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.090	0.097	0.028	0.033	0.130	0.138	0.104	0.110
N	391831	391797	392310	392276	301404	301379	119511	119505

Notes. This table reports the effect of the number of licensed premises on four alcohol-related outcomes. Columns (1) and (2) shows the effect on the normalized drinking frequency (6-point scale), Columns (3) and (4) on the normalized probability of drinking daily, Columns (5) and (6) on the normalized log. number of alcoholic beverages consumed per week, and Columns (7) and (8) on the normalized amount of alcohol drunk in a typical drinking session. “On-premise licenses (per 1,000)” indicates the number of premises per capita allowed to sell alcohol on-premise (including premises permitted to sell alcohol both on- and off-premise). “PGS-DPW” denotes the polygenic score for the number of drinks per week. “On-premise licenses (per 1,000) \times PGS-DPW” is the interaction of the policy with the polygenic score. All models control for genetic stratification using the first 40 principal components of the genetic data and “Age \times Sex” fixed effects. Standard errors are clustered at the local authority level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

high genetic propensity to drink react much less to the policy. A one standard deviation increase in the PGS reduces the effect of the policy in by roughly a quarter for the drinking frequency ($P = 0.002$), daily drinking ($P = 0.002$), and the number of drinks per week ($P = 0.001$), and we find a negative and significant interaction for the typical amount drunk ($P = 0.000$). Figure A4.6 in Appendix A4.I displays the relationship between the number of licensed premises and the four outcomes for UKB participants above and below the median PGS.

As licensing committees also control other forms of entertainment, we can use the number of no-sales licenses—premises allowed to offer licensable entertainment, but not to sell alcohol—as a measure for licensing policy explicitly aimed at reducing alcohol sales. Using the number of no-sales licenses per 1,000 residents, we find that more restrictive licensing is associated with a decrease in alcohol consumption for all outcomes (Table 4.3). Residents drink less frequently ($P = 0.014$) and fewer report that they drink daily ($P = 0.007$). They also report that they consume fewer alcoholic beverages a week ($P = 0.009$) and that they drink less in a typical drinking session ($P = 0.005$). The interaction models again show that UKB participants with a higher genetic propensity drink significantly more and react less to the policy. A one standard deviation increase in the PGS roughly halves the effect on the drinking frequency ($P = 0.002$), reduces the change in daily drinking by about a quarter ($P = 0.042$), and cuts the effect on the amount of alcohol consumed by a third for the weekly amount ($P = 0.056$) and the typical amount, although not significantly ($P = 0.147$). Figure A4.7 in Appendix A4.I displays the relationship between the number of no-sales licenses and the four outcomes for UKB participants

TABLE 4.3
GENE-ENVIRONMENT INTERACTION FOR LICENSING POLICY (NO-SALES LICENSES)

	Drinking frequency		Daily drinker		Log. drinks per week		Typical amount	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
No-sales premise licenses (per 1,000)	-0.809*	-0.764*	-0.841**	-0.812**	-0.596*	-0.571**	-0.638**	-0.596**
	(0.328)	(0.321)	(0.303)	(0.301)	(0.234)	(0.216)	(0.215)	(0.213)
PGS-DPW		0.136***		0.112***		0.151***		0.125***
		(0.007)		(0.008)		(0.007)		(0.011)
No-sales premise licenses (per 1,000) \times PGS-DPW		0.345**		0.214*		0.210 ⁺		0.224
		(0.110)		(0.105)		(0.109)		(0.154)
Constant	0.045 ⁺	0.044 ⁺	0.057*	0.056*	0.023	0.020	0.027 ⁺	0.025 ⁺
	(0.025)	(0.025)	(0.025)	(0.025)	(0.015)	(0.015)	(0.014)	(0.014)
First 40 PCs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age \times Sex	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.089	0.097	0.027	0.032	0.129	0.137	0.105	0.111
N	385443	385409	385900	385866	296467	296442	122009	122003

Notes. This table reports the effect of the number of premises allowed to provide licensable entertainment but not permitted to sell alcohol on four alcohol-related outcome. Columns (1) and (2) shows the effect on the normalized drinking frequency (6-point scale), Columns (3) and (4) on the normalized probability of drinking daily, Columns (5) and (6) on the normalized log. number of alcoholic beverages consumed per week, and Columns (7) and (8) on the normalized amount of alcohol drunk in a typical drinking session. “No-sales premise licenses (per 1,000)” indicates the number of premises per capita allowed to provide licensable entertainment but not permitted to sell alcohol. “PGS-DPW” denotes the polygenic score for the number of drinks per week. “No-sales premise licenses (per 1,000) \times PGS-DPW” is the interaction of the policy with the polygenic score. All models control for genetic stratification using the first 40 principal components of the genetic data and “Age \times Sex” fixed effects. Standard errors are clustered at the local authority level. Significance levels: ⁺ $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

above and below the median PGS.

In addition to controlling the number of alcohol sales points, the Licensing Act relaxed the traditional 11 a.m. to 11 p.m. window for alcohol sales and granted licensing committees the power to restrict or extend the hours that pubs and supermarkets are allowed to sell alcohol. While we do not have sufficient data on pub opening hours to test directly for the impact of this policy, we do have a proxy in the form of 24h sales licenses for supermarkets (Table 4.4). Unlike in the previous analysis, the evidence on the direct effect of the policy is somewhat mixed. While the estimates point in the expected direction, only the proportion of daily drinkers changes significantly ($P = 0.008$). The results of the interaction model are nevertheless consistent with the results from the other policy measures: we find significant and seizable negative interactions for the drinking frequency ($P = 0.024$), daily drinking ($P = 0.019$), and the number of drinks per week ($P = 0.020$). Figure A4.8 in Appendix A4.I displays the relationship between the number of 24h licenses and the four outcomes for UKB participants above and below the median PGS.

While the interaction between the number of pubs and retailers and the PGS might be strongly affected by participants’ self-selection into particular areas, this is less of a concern for licensing policy, which is measured at the local authority level. Indeed, while an individual’s genetic propensity to drink is predictive of movement within local authorities, we do not find evidence that the decision to move between local authorities is affected by the fit between participants’ genetic predisposition and the local licensing policy (see Table A4.7 in Appendix A4.II). Moreover, as licensing committees are not allowed to determine their policy to advance public health objectives—much less in response to citizens’ genetic endowments—, these results

TABLE 4.4
GENE-ENVIRONMENT INTERACTION FOR LICENSING POLICY (24H SUPERMARKETS)

	Drinking frequency		Daily drinker		Log. drinks per week		Typical amount	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
24h supermarket licenses (per 1,000)	4.268 (3.595)	4.470 (3.630)	8.712** (3.270)	8.824** (3.315)	4.084 (2.727)	4.143 (2.734)	1.698 (1.998)	1.435 (1.927)
PGS-DPW		0.162*** (0.003)		0.131*** (0.004)		0.167*** (0.003)		0.141*** (0.006)
24h supermarket licenses (per 1,000) \times PGS-DPW		-2.138* (0.943)		-1.494* (0.635)		-1.827* (0.781)		-0.486 (0.839)
Constant	-0.007 (0.014)	-0.007 (0.015)	-0.007 (0.014)	-0.006 (0.014)	-0.018 ⁺ (0.009)	-0.020* (0.009)	-0.018* (0.009)	-0.016 ⁺ (0.009)
First 40 PCs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age \times Sex	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.088	0.095	0.027	0.032	0.127	0.135	0.105	0.111
N	418654	418619	419158	419123	322719	322694	123743	123738

Notes. This table reports the effect of the number of supermarkets licensed to sell alcohol 24h a day on four alcohol-related outcomes. Columns (1) and (2) shows the effect on the normalized drinking frequency (6-point scale), Columns (3) and (4) on the normalized probability of drinking daily, Columns (5) and (6) on the normalized log. number of alcoholic beverages consumed per week, and Columns (7) and (8) on the normalized amount of alcohol drunk in a typical drinking session. “24h supermarket licenses (per 1,000)” indicates the number of supermarkets per capita licensed to sell alcohol 24h a day. “PGS-DPW” denotes the polygenic score for the number of drinks per week. “24h supermarket licenses (per 1,000) \times PGS-DPW” is the interaction of the policy with the polygenic score. All models control for genetic stratification using the first 40 principal components of the genetic data and “Age \times Sex” fixed effects. Standard errors are clustered at the local authority level. Significance levels: ⁺ $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

are strong evidence that the individuals with the highest propensity to drink react least to supply-focused alcohol policy.

E. Genetic Predisposition and Public Health

Although there is no truly safe level of alcohol consumption (Department of Health, 2016; Burton and Sheron, 2018), a disproportionate share of the public health burden is driven by those individuals with the heaviest drinking behavior (Burton et al., 2017). In the final part of our analysis, we show that the polygenic score can identify UKB participants with the highest risk of developing an alcohol-related disease, even if we control for self-reported alcohol consumption.

Being primarily designed to study the contribution of genetic predisposition and exposure to environmental risk factors to disease in late adulthood, the UKB encodes comprehensive information from participants’ NHS health records, including high-quality data on physician-diagnosed medical conditions. These conditions are coded according to the ICD-10 classification system (WHO, 2007), allowing us to test explicitly for the effect on alcohol-related diseases. We restrict our analysis to conditions described as alcohol-related or alcohol-induced in the ICD-10 with at least 25 diagnosed cases in the UKB.

We find a significant positive relationship between the PGS and alcohol-related medical conditions of the liver (see Table A4.8 in Appendix A4.II). Relative to the mean incident rate, a one standard deviation increase in the polygenic score is associated with a +40.3% rise ($P = 0.000$) in the prevalence of alcohol-related liver diseases (ICD-10 K70). This association persists even

if we control for participants' self-reported drinking frequency (+42.3%, $P = 0.000$). This holds for each of the three most common alcohol-related liver conditions in the UKB—alcoholic hepatitis (ICD-10 K70.1, +32.8%, $P = 0.027$), alcoholic liver cirrhosis (ICD-10 K70.3, +26.2%, $P = 0.033$), and alcoholic liver failure (K70.4, +36.4%, $P = 0.031$)—as well as for other unspecified alcohol-related liver diseases (ICD-10 K70.9, +54.0%, $P = 0.000$), but not for alcoholic fatty liver (ICD-10 K70.0, +68.0%, $P = 0.115$).

While afflictions of the liver are the most widely known consequences of excessive drinking, alcohol also has significant long-run effects on mental health. Repeating our analysis for psychological disorders, we find that the PGS is significantly associated with heightened rates of alcohol-related mental health conditions (see Table A4.9 in Appendix A4.II). A one standard deviation increase in the polygenic score is associated with a +39.6% rise ($P = 0.000$) in mental and behavioral disorders due to use of alcohol (ICD-10 F10). Again, this relationship remains quantitatively similar and significant if we control for self-reported alcohol intake (+40.4%, $P = 0.000$). A more detailed look reveals that the effect is largely driven by the most common conditions: acute intoxication (ICD-10 F10.0, +29.9%, $P = 0.001$), harmful use (ICD-10 F10.1, +33.1%, $P = 0.027$), alcohol dependence syndrome (ICD-10 F10.2, +54.8%, $P = 0.000$) and alcohol withdrawal state (ICD-10 F10.3 and F10.4²⁰, +49.3%, $P = 0.000$), there is no significant association with the much rarer alcoholic psychotic disorder (ICD-10 F10.5, +46.4%, $P = 0.144$) or other unspecified alcohol-related mental disorders (ICD-10 F109, +58.4%, $P = 0.179$).

Finally, we tested the relationship with a number of alcohol-related afflictions in other ICD-10 code classes (see Table A4.10 in Appendix A4.II). Again controlling for self-reported drinking, we find significant associations for alcoholic gastritis (ICD-10 K29.2, +56.1%, $P = 0.003$) and alcohol-induced chronic pancreatitis (ICD-10 K86.0, +63.4%, $P = 0.000$, although not for alcohol-induced acute pancreatitis (ICD-10 K85.2, +42.2%, $P = 0.245$) or alcoholic degeneration of the nervous system (ICD-10 G31.2, +49.2%, $P = 0.145$).

What accounts for the predictive power of the polygenic score on top of self-reported drinking behavior? While the polygenic score may capture genetic markers that directly contribute to the pathogenesis of these health conditions, it is more likely that the polygenic score is a more accurate measure of life-time alcohol intake than the consumption measures in the UKB. This is partly by design—participants are asked about their current drinking behavior—but we cannot rule out that it is also driven by the self-reported nature of the measures, as people tend to underestimate their alcohol consumption (Poikolainen, 1985; Gmel and Rehm, 2004; Gual et al., 2017).

Taken together, our final results show that individuals with a high PGS are significantly more likely to have diagnosed alcohol-related medical conditions, including liver diseases, psychological disorders, and various afflictions of the digestive system. Moreover, these results hold even if we control for self-reported alcohol consumption, illustrating how incorporating genetic

20. Withdrawal state without and with delirium.

information can help us uncover relevant dimensions of health-inequality that are not commonly observable.

IV. CONCLUSION

Overall, our results demonstrate how genetic information can shed light on the determinants and the dynamics of health inequalities and how genetic endowments interact with individual choices and public health policy. We show that the effectiveness of a supply-focused licensing policy as a tool to mitigate alcohol abuse can clash with individual predispositions and might actually exacerbate inherited inequality.

By combining genetic data with fine-grained geographical measures of alcohol availability, we show that individuals with a high genetic propensity to drink self-select into environments with easier access to alcohol, react less to changes in the availability of alcohol, and respond less to restrictive licensing. As the same people also account for a disproportionate share of the alcohol-related public health burden, our results suggest that supply-focused alcohol policies are less effective precisely for the people who might need them the most. Demand-side policies that focus on an individual's willingness to engage in unhealthy behaviors—such as nudges, support groups, or individual therapies—might be more effective at curtailing health inequality and helping those more predisposed for such behaviors.

Our work suffers from some limitations. The need for geographic variation restricts our analysis to policy determined at the local level in England and Wales. This excludes some advocated policy tools such as higher alcohol taxes (Cook, 1982; Anderson and Baumberg, 2006; Wagenaar et al., 2010) or additional taxes on selected alcohol beverages (Müller et al., 2010), business-led initiatives such as the “Billion Units Pledge” (Department of Health, 2011; Gornall, 2014; Knai et al., 2015), minimum unit prices (Purshouse et al., 2010; Brennan et al., 2014; Holmes et al., 2014), as well as changes in the legal drinking age (Toomey et al., 1996; Carpenter and Dobkin, 2011) and their enforcement (Wagenaar and Wolfson, 1994; Retail of Alcohol Standards Group, 2014), and bans on promotions (Anderson, 2009). It remains an open question whether these other policy measures can mitigate the effects of genetic inequality instead of compounding them.

Finally, our finding that the polygenic score predicts alcohol-related disease, even after controlling for self-reported alcohol consumption, suggests a more targeted approach as a possible remedy. Future research should explore the use of genetic information to identify individuals at risk of alcohol dependence and to provide targeted interventions to address health-inequality.

REFERENCES

ABDELLAOUI, A., D. HUGH-JONES, K. E. KEMPER, Y. HOLTZ, M. G. NIVARD, L. VEUL, L. YENGO, B. P. ZIETSCH, T. M. FRAYLING, N. R. WRAY, J. YANG, K. J. H. VERWEIJ, AND P. M. VISSCHER

- (2018): “Genetic Consequences of Social Stratification in Great Britain,” *bioRxiv*, 457515.
- ANDERSON, P. (2009): “Is it time to ban alcohol advertising?” *Clinical Medicine*, 9, 121–124.
- ANDERSON, P. AND B. BAUMBERG (2006): *Alcohol in Europe—a public health perspective.*, European Commission.
- BANZHAF, H. S. AND R. P. WALSH (2008): “Do people vote with their feet? An empirical test of Tiebout,” *American Economic Review*, 98, 843–63.
- BARCELLOS, S. H., L. S. CARVALHO, AND P. TURLEY (2018): “Education can reduce health differences related to genetic risk of obesity.” *Proceedings of the National Academy of Sciences*, 201802909.
- BBPA (2016): *Statistical Handbook 2016*, London: British Beer and Pub Association.
- BELSKY, D. W., A. CASPI, L. ARSENEAULT, D. L. CORCORAN, B. W. DOMINGUE, K. M. HARRIS, R. M. HOUTS, J. S. MILL, T. E. MOFFITT, J. PRINZ, K. SUGDEN, J. WERTZ, B. WILLIAMS, AND C. L. ODGERS (2019): “Genetics and the geography of health, behaviour and attainment,” *Nature Human Behaviour*, 1.
- BELSKY, D. W., B. W. DOMINGUE, R. WEDOW, L. ARSENEAULT, AND J. D. BOARDMAN (2018): “Genetic analysis of social-class mobility in five longitudinal studies,” *Proceedings of the National Academy of Sciences*, 1–10.
- BENJAMIN, D. J., D. CESARINI, D. I. LAIBSON, AND P. TURLEY (Forthcoming): “Social Science Genetics: A Primer and Progress Report,” *Journal of Economic Literature*.
- BENJAMINI, Y. AND Y. HOCHBERG (1995): “Controlling the false discovery rate: A practical and powerful approach to multiple testing,” *Journal of the Royal Statistical Society. Series B*, 57, 289–300.
- BENJAMINI, Y. AND D. YEKUTIELI (2005): “False discovery rate-adjusted multiple confidence intervals for selected parameters,” *Journal of the American Statistical Association*, 100, 71–81.
- BOWLES, S. (1970): “Migration as investment: Empirical tests of the human investment approach to geographical mobility,” *The Review of Economics and Statistics*, 356–362.
- BRAUDT, D. B. (2018): “Sociogenomics in the 21st century: An introduction to the history and potential of genetically informed social science,” *Sociology Compass*, 12, 1–16.
- BRENNAN, A., Y. MENG, J. HOLMES, D. HILL-MCMANUS, AND P. S. MEIER (2014): “Potential benefits of minimum unit pricing for alcohol versus a ban on below cost selling in England 2014: modelling study,” *BMJ*, 349, g5452.
- BROWN, L. A. AND E. G. MOORE (1970): “The intra-urban migration process: a perspective,” *Geografiska Annaler: Series B, Human Geography*, 52, 1–13.
- BURTON, R., C. HENN, D. LAVOIE, R. O’CONNOR, C. PERKINS, K. SWEENEY, F. GREAVES, B. FERGUSON, C. BEYNON, A. BELLONI, ET AL. (2017): “A rapid evidence review of the effectiveness and cost-effectiveness of alcohol control policies: an English perspective,” *The Lancet*, 389, 1558–1580.
- BURTON, R. AND N. SHERON (2018): “No level of alcohol consumption improves health,” *The Lancet*, 392, 987–988.
- CARPENTER, C. AND C. DOBKIN (2011): “The minimum legal drinking age and public health,” *Journal of Economic Perspectives*, 25, 133–56.
- COLLINS, R. (2012): “What makes UK Biobank special?” *The Lancet*, 379, 1173–1174.

- COOK, P. J. (1982): "Alcohol taxes as a public health measure," *British Journal of Addiction*, 77, 245–250.
- DARWIN, C. (1859): *The Origin of Species; And, the Descent of Man*, Modern library.
- DEPARTMENT OF HEALTH (2011): "The public health responsibility deal," .
- (2016): "UK Chief Medical Officers' Low Risk Drinking Guidelines," Tech. Rep. August, Department of Health, London.
- DUDBRIDGE, F. (2013): "Power and predictive accuracy of polygenic risk scores," *PLoS genetics*, 9, e1003348.
- FREESE, J. (2018): "The Arrival of Social Science Genomics," *Contemporary Sociology: A Journal of Reviews*, 47, 524–536.
- FREUD, S. (1930): *Civilization and its discontents*, WW Norton & Company.
- GMEL, G. AND J. REHM (2004): "Measuring alcohol consumption," *Contemporary Drug Problems*, 31, 467–540.
- GORNALL, J. (2014): "Is the billion unit pledge just window dressing?" *BMJ*, 348, g3190.
- GRANT, B. F., D. A. DAWSON, F. S. STINSON, S. P. CHOU, M. C. DUFOUR, AND R. P. PICKERING (2004): "The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991–1992 and 2001–2002," *Drug and alcohol dependence*, 74, 223–234.
- GUAL, A., J. ÁNGEL ARBESÚ, J. ZARCO, M. D. L. M. BALCELLS-OLIVERÓ, H. LÓPEZ-PELAYO, L. MIQUEL, AND J. BOBES (2017): "Risky drinkers underestimate their own alcohol consumption," *Alcohol and Alcoholism*, 52, 516–517.
- HARDEN, K. P., J. E. HILL, E. TURKHEIMER, AND R. E. EMERY (2008): "Gene-environment correlation and interaction in peer effects on adolescent alcohol and tobacco use," *Behavior genetics*, 38, 339–347.
- HAWORTH, S., R. MITCHELL, L. CORBIN, K. H. WADE, T. DUDDING, A. BUDU-AGGREY, D. CARSLAKE, G. HEMANI, L. PATERNOSTER, G. D. SMITH, N. DAVIES, D. LAWSON, AND N. TIMPSON (2018): "Common genetic variants and health outcomes appear geographically structured in the UK Biobank sample: Old concerns returning and their implications." *bioRxiv*, 294876.
- HECKMAN, J. J. (2007): "The economics, technology, and neuroscience of human capability formation." *Proceedings of the National Academy of Sciences*, 104, 13250–5.
- HENSEN, M. M., M. R. DE VRIES, AND F. CÖRVERS (2009): "The role of geographic mobility in reducing education-job mismatches in the Netherlands," *Papers in Regional Science*, 88, 667–682.
- HOLMES, J., Y. MENG, P. S. MEIER, A. BRENNAN, C. ANGUS, A. CAMPBELL-BURTON, Y. GUO, D. HILL-MCMANUS, AND R. C. PURSHOUSE (2014): "Effects of minimum unit pricing for alcohol on different income and socioeconomic groups: a modelling study," *The Lancet*, 383, 1655–1664.
- HOME OFFICE (2012): *Amended guidance issued under Section 182 of the Licensing Act 2003*, The Stationery Office, ISBN: 9780108511400.
- HUME, D. (1748): *Philosophical essays concerning human understanding*, A. Millar.
- HUNTER, D. J. (2005): "Gene-environment interactions in human diseases." *Nature Reviews Genetics*, 6, 287–98.

- KIDD, M. P., N. O'LEARY, AND P. SLOANE (2017): "The impact of mobility on early career earnings: A quantile regression approach for UK graduates," *Economic Modelling*, 62, 90–102.
- KNAI, C., M. PETTICREW, M. A. DURAND, C. SCOTT, L. JAMES, A. MEHROTRA, E. EASTMURE, AND N. MAYS (2015): "The Public Health Responsibility deal: has a public-private partnership brought about action on alcohol reduction?" *Addiction*, 110, 1217–1225.
- LIU, M., Y. JIANG, R. WEDOW, Y. LI, D. M. BRAZEL, F. CHEN, G. DATTA, J. DAVILA-VELDERRAIN, D. MCGUIRE, C. TIAN, X. ZHAN, H. CHOQUET, A. R. DOCHERTY, J. D. FAUL, J. R. FOERSTER, L. G. FRITSCH, M. E. GABRIELSEN, S. D. GORDON, J. HAESSLER, J.-J. HOTTENGA, H. HUANG, S.-K. JANG, P. R. JANSEN, Y. LING, R. MÄGI, N. MATOBA, G. MCMAHON, A. MULAS, V. ORRÛ, T. PALVIAINEN, A. PANDIT, G. W. REGINSSON, A. H. SKOGHOLT, J. A. SMITH, A. E. TAYLOR, C. TURMAN, G. WILLEMSSEN, H. YOUNG, K. A. YOUNG, G. J. M. ZAJAC, W. ZHAO, W. ZHOU, G. BJORNSDOTTIR, J. D. BOARDMAN, M. BOEHNKE, D. I. BOOMSMA, C. CHEN, F. CUCCA, G. E. DAVIES, C. B. EATON, M. A. EHRINGER, T. ESKO, E. FIORILLO, N. A. GILLESPIE, D. F. GUDBJARTSSON, T. HALLER, K. M. HARRIS, A. C. HEATH, J. K. HEWITT, I. B. HICKIE, J. E. HOKANSON, C. J. HOPFER, D. J. HUNTER, W. G. IACONO, E. O. JOHNSON, Y. KAMATANI, S. L. R. KARDIA, M. C. KELLER, M. KELLIS, C. KOOPERBERG, P. KRAFT, K. S. KRAUTER, M. LAAKSO, P. A. LIND, A. LOUKOLA, S. M. LUTZ, P. A. F. MADDEN, N. G. MARTIN, M. MCGUE, M. B. MCQUEEN, S. E. MEDLAND, A. METSPALU, K. L. MOHLKE, J. B. NIELSEN, Y. OKADA, U. PETERS, T. J. C. POLDERMAN, D. POSTHUMA, A. P. REINER, J. P. RICE, E. RIMM, R. J. ROSE, V. RUNARSDOTTIR, M. C. STALLINGS, A. STANČÁKOVÁ, H. STEFANSSON, K. K. THAI, H. A. TINDLE, T. TYRFINGSSON, T. L. WALL, D. R. WEIR, C. WEISNER, J. B. WHITFIELD, B. S. WINSVOLD, J. YIN, L. ZUCCOLO, L. J. BIERUT, K. HVEEM, J. J. LEE, M. R. MUNAFÒ, N. L. SACCONE, C. J. WILLER, M. C. CORNELIS, S. P. DAVID, D. A. HINDS, E. JORGENSEN, J. KAPRIO, J. A. STITZEL, K. STEFANSSON, T. E. THORGEIRSSON, G. ABECASIS, D. J. LIU, AND S. VRIEZE (2019): "Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use," *Nature Genetics*.
- MACHIELA, M. J., C.-Y. CHEN, C. CHEN, S. J. CHANOCK, D. J. HUNTER, AND P. KRAFT (2011): "Evaluation of polygenic risk scores for predicting breast and prostate cancer risk," *Genetic epidemiology*, 35, 506–514.
- MAK, T., R. M. PORSCH, S. W. CHOI, AND P. C. SHAM (2018): "Polygenic scores for UK Biobank scale data," *bioRxiv*, 252270.
- MARTINEAU, F., H. GRAFF, C. MITCHELL, AND K. LOCK (2013): "Responsibility without legal authority? Tackling alcohol-related health harms through licensing and planning policy in local government," *Journal of Public Health*, 36, 435–442.
- MOKDAD, A. H., J. S. MARKS, D. F. STROUP, AND J. L. GERBERDING (2004): "Actual causes of death in the United States, 2000." *JAMA*, 291, 1238–45.
- MULCASTER, R. (1582): *Mulcaster's Elementaire*, London: Clarendon Press.
- MÜLLER, S., D. PIONTEK, A. PABST, S. E. BAUMEISTER, AND L. KRAUS (2010): "Changes in alcohol consumption and beverage preference among adolescents after the introduction of the alcopops tax in Germany," *Addiction*, 105, 1205–1213.
- NECHYBA, T. J. (2000): "Mobility, targeting, and private-school vouchers," *American Economic Review*, 90, 130–146.
- ONS (2018): "Pub closures: Response to Freedom of Information Request," <https://www.ons.gov.uk/aboutus/transparencyandgovernance/freedomofinformationfoi/pubclosures>., accessed on December 27, 2019.

- PEI, Z., J.-S. PISCHKE, AND H. SCHWANDT (2019): “Poorly measured confounders are more useful on the left than on the right,” *Journal of Business & Economic Statistics*, 37, 205–216.
- PLOMIN, R., J. C. DEFRIES, AND J. C. LOEHLIN (1977): “Genotype-environment interaction and correlation in the analysis of human behavior,” *Psychological Bulletin*, 84, 309–22.
- POIKOLAINEN, K. (1985): “Underestimation of recalled alcohol intake in relation to actual consumption,” *British journal of addiction*, 80, 215–216.
- PRICE, A., N. PATTERSON, R. PLENGE, M. WEINBLATT, N. SHADICK, AND D. REICH (2006): “Principal components analysis corrects for stratification in genome-wide association studies,” *Nature Genetics*, 38, 904.
- PURSHOUSE, R. C., P. S. MEIER, A. BRENNAN, K. B. TAYLOR, AND R. RAFIA (2010): “Estimated effect of alcohol pricing policies on health and health economic outcomes in England: an epidemiological model,” *The Lancet*, 375, 1355–1364.
- RETAIL OF ALCOHOL STANDARDS GROUP (2014): “Rising to the Challenge: A report into the application and impact of Challenge 25,” .
- RIETVELD, C. A., D. C. CONLEY, N. ERIKSSON, T. ESKO, S. E. MEDLAND, A. A. E. VINKHUYZEN, J. YANG, J. D. BOARDMAN, C. F. CHABRIS, C. T. DAWES, B. W. DOMINGUE, D. A. HINDS, M. JOHANNESSON, A. K. KIEFER, D. I. LAIBSON, P. K. E. MAGNUSSON, J. L. MOUNTAIN, S. OSKARSSON, O. ROSTAPSHOVA, A. TEUMER, J. TUNG, P. M. VISSCHER, D. J. BENJAMIN, D. CESARINI, AND P. D. KOELLINGER (2014): “Replicability and Robustness of Genome-Wide Association Studies for Behavioral Traits,” *Psychological Science*, 25, 1975–1986.
- ROOM, R., T. BABOR, AND J. REHM (2005): “Alcohol and public health,” *The Lancet*, 365, 519–530.
- SCHMITZ, L. L. AND D. C. CONLEY (2016a): “The Impact of Late-Career Job Loss and Genotype on Body Mass Index,” *NBER Working Paper*.
- (2016b): “The Long-Term Consequences of Vietnam-Era Conscript and Genotype on Smoking Behavior and Health,” *Behavior Genetics*, 46, 43–58.
- (2017): “The effect of Vietnam-era conscription and genetic potential for educational attainment on schooling outcomes,” *Economics of Education Review*, 61, 85–97.
- SCOTTISH PARLIAMENT (2005): “Licensing (Scotland) Act 2005,” *Scottish Parliament: Edinburgh, UK*.
- SIMMONS, J. W. (1968): “Changing residence in the city: a review of intraurban mobility,” *Geographical Review*, 622–651.
- SUDLOW, C., J. GALLACHER, N. ALLEN, V. BERAL, P. BURTON, J. DANESH, P. DOWNEY, P. ELLIOTT, J. GREEN, M. LANDRAY, ET AL. (2015): “UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age,” *PLoS medicine*, 12, e1001779.
- TOOMEY, T. L., C. ROSENFELD, AND A. C. WAGENAAR (1996): “The minimum legal drinking age,” *Alcohol Research*, 20, 213.
- UKB (2006): “UK Biobank: Protocol for a large-scale prospective epidemiological resource,” *Protocol No: UKBB-PROT-09-06 (Main Phase)*.
- VILHJÁLMSSON, B. J., J. YANG, H. K. FINUCANE, A. GUSEV, S. LINDSTRÖM, S. RIPKE, G. GENOVESE, P.-R. LOH, G. BHATIA, R. DO, ET AL. (2015): “Modeling linkage disequilibrium increases accuracy of polygenic risk scores,” *The American Journal of Human Genetics*, 97, 576–592.

- WAGENAAR, A. C., A. L. TOBLER, AND K. A. KOMRO (2010): “Effects of alcohol tax and price policies on morbidity and mortality: a systematic review,” *American journal of public health*, 100, 2270–2278.
- WAGENAAR, A. C. AND M. WOLFSON (1994): “Enforcement of the legal minimum drinking age in the United States,” *Journal of public health policy*, 15, 37–53.
- WHO (2007): *International classification of diseases and related health problems, 10th revision.*, Geneva, Switzerland: World Health Organization.
- WILSNACK, R. W., N. D. VOGELTANZ, S. C. WILSNACK, AND T. R. HARRIS (2000): “Gender differences in alcohol consumption and adverse drinking consequences: cross-cultural patterns,” *Addiction*, 95, 251–265.
- WILSNACK, R. W., S. C. WILSNACK, A. F. KRISTJANSON, N. D. VOGELTANZ-HOLM, AND G. GMEL (2009): “Gender and alcohol consumption: patterns from the multinational GENACIS project,” *Addiction*, 104, 1487–1500.
- YANKOW, J. J. (2003): “Migration, job change, and wage growth: a new perspective on the pecuniary return to geographic mobility,” *Journal of Regional Science*, 43, 483–516.

Appendix A2

Appendices to: Civic Honesty around the Globe

I. MATERIALS AND METHODS

A. *Lost Wallet Experiments*

We visited 355 cities in 40 countries and turned in a total of 17,303 wallets between July 2013 and December 2016. Table A2.1 provides an overview of the study design, including the countries and cities covered, the amount of money included in the wallets, the names on the business cards, the items on the shopping list, and the number of observations. Our study was approved by the Human Subjects Committee of the Faculty of Economics, Business Administration, and Information Technology at the University of Zurich.

B. *Selection of Countries and Cities*

We selected our sample of countries based on several factors, most important being that the country have a sufficient number of large cities. As a rough guide during the planning process we aimed for populations of at least 100,000, but used this rule flexibly as availability of feasible drop-off locations varied substantially even when restricting ourselves to large cities. In addition to city size, a country had to be relatively easy to visit and safe enough for our research assistants to perform the wallet drop-offs. Customs, immigration, and banking regulation also played a role because research assistants needed to either import or withdraw sufficient money to place in the wallets.

For each country we typically chose five to eight cities to perform the wallet drop-offs. We took the largest cities of a country as a starting point and adapted the list to accommodate safety concerns, cover the main regions of a country, and avoided cities that belong to the same metropolitan area. As cities differed in their size, the number of drop-offs in a city was determined by the relative population size using the following formula:

$$N_i = \frac{\sqrt{POP_i}}{\sum_{k \in C} \sqrt{POP_k}} * N_C^{target}, \quad (\text{A2.1})$$

where N_i is the number of drop-offs in city i , POP_i is city i 's population size, k is a city sampled from country C , and N_C^{target} is the target sample size for a given country. This adjustment was designed to avoid a single city dominating our estimates of a country's response rate, while also giving greater weight to more populated cities as they represent a greater fraction of a country's total population (and so tend to be more influential politically, culturally, and economically).

C. *Number of Drop-Offs*

We usually collected 400 observations per country, but there were exceptions to this rule. For some countries we set a different target sample size, and for other countries we ended up

with deviations from the targeted sample size due to unforeseen circumstances or minor errors in the data collection process.

We collected a greater number of observations in the US, UK, and Poland since we ran two additional treatment conditions (BigMoney and Money-NoKey conditions) in these countries.¹ In the United States, we collected 300 wallets each in the NoMoney and Money conditions and 200 wallets each in the Money-NoKey and BigMoney conditions, yielding a total sample of 1,000 observations. In the UK, we turned in 200 wallets each in the NoMoney, BigMoney, and Money-NoKey condition, and 600 wallets in the Money condition. We were unable to track email responses for 67 wallets in the Money condition and one wallet in the BigMoney condition due to a procedural error, leaving us with a total of 1,132 observations in the UK. In Poland, we turned in 200 wallets in each of our four conditions, yielding a total sample of 800 observations.

For eight countries—Croatia, Denmark, Ghana, Israel, Kenya, Norway, Serbia, and Russia—we set a sample size target of 300 drop-offs due to either a limited number of sufficiently large cities or due to safety concerns. For India, we made a last minute change by replacing Chennai with Coimbatore due to severe flooding that took place in February 2015. In Kenya we did not carry out data collection in the last city visited (Malindi) because the research assistant was arrested and interrogated by the military police for suspicious activity. In Chile, four wallets had to be excluded from the analysis because of a handling mistake which made it impossible to ascertain the location of where the wallets were turned in.

Additional minor deviations from the target sample size occurred due to rounding errors in the allocation of drop-offs to different cities or because experimenters could not find a suitable replacement for a closed drop-off location in time. Countries with minor deviations are marked by a footnote in Table A2.1.

D. Selection of Drop-Off Locations

We focused on five types of institutions as drop-off locations: (i) banks, (ii) theaters, museums, or other cultural establishments, (iii) post offices, (iv) hotels, and (v) police stations, courts of law, and other public offices. While we aimed at an equal distribution of institutions, this was not always feasible. In particular, post offices were sometimes hard to find near city centers as they are often spread over geographic regions. Our final distribution was 23% for banks, 20% for cultural establishments, 14% for post offices, 22% for hotels, and 21% for public offices.

Drop-off locations were always planned in advance. To find appropriate locations, we used

1. In the US, UK, Poland, France, Italy, and Spain we also conducted additional treatment arms which were orthogonal to our NoMoney and Money conditions. These additional treatment arms mostly involved changing subtle characteristics about the owner of the wallet. We plan to report these results in a separate paper. For France and the UK we observed no significant effect on the reporting rate across these additional conditions, so we pool the data for those two countries here to increase the precision of our estimates. Excluding this additional data from the analysis has virtually no effect on the results we report below.

official websites (e.g., for police stations), travel guides (e.g., for hotels and museums), and Google Maps. To reduce travel time, we advised research assistants to select drop-off locations close to a city center and to choose drop-off locations within walking distance of each other. To avoid suspicion, we excluded drop-off locations that were next to or across the street from one another. We also advised research assistants to select locations that were far enough away from a given police station as to reduce the risk that multiple recipients would turn in wallets to the same police station. When available, we used Google Street View to verify that a location still existed and that the location was easily accessible from the street. Prior to performing the drop-offs, research assistants also checked for national and local holidays, opening hours, and specific working culture (e.g., siesta in Spain).

E. The Wallets

Our wallets were transparent business card cases (see Figure A2.1). We used transparent cases to ensure that recipients could inspect the wallet’s contents without having to open it. Each wallet contained the same personal items: (i) three identical business cards, (ii) a grocery list, and (iii) a key. The business cards displayed the owner’s name, email address, and job title. Their purpose was to identify the owner and provide contact details.

The business cards and shopping list serve to identify the owner as a local resident, signaling that it would be relatively easy to contact the owner and return the wallet. For the business cards, we typically created three fictitious male owners for each country using common local names. We used several sources to assemble lists of common first and last names, which we then checked to avoid names used as references for generic or unidentified persons (e.g., John Doe), were shared with celebrities, or led to a single user-profile on Facebook. The business cards provided the owner’s email address, and identified him as a freelance software engineer (to avoid attempts by recipients to reach the owner through his place of employment).

There were some exceptions to how we generated business cards and shopping lists for our wallets. In Switzerland and the Czech Republic, we used the real name of research assistants so that we would be able to collect reported wallets for our internal validation check. For these two countries we also decided to use only two identities (rather than three) so that we could pick up a larger share of the wallets. In Canada and India, different names were used for some cities to accommodate for the local language. Due to South Africa’s history of race relations, we used two discernibly white and two discernibly black names, leading to a total of four names.² We made occasional changes to the shopping lists to accommodate local customs, such as using rice instead of pasta or substituting milk with some other beverage where lactose intolerance was common. Table A2.1 provides a comprehensive list of names and shopping lists.

2. In South Africa reporting rates were remarkably similar between Black and White names. Reporting rates were always between 32% and 35%, with no significant difference in reporting rates between the four identities ($\chi^2_3 = 0.255$, $P = 0.968$).



FIGURE A2.1
Example Lost Wallet

Notes. Example of a wallet used in our field experiments. All wallets belonged to a male software developer with country-specific names (see Table A2.1 for the complete list of names). We placed the business cards in the wallets so that this information was visible to all participants. The wallet dimensions were 93mm x 59mm x 5mm and it weighed approximately 24 grams in the NoMoney condition.

F. Drop-Off Procedure

We recruited eleven male and two female research assistants to perform the drop-offs. All research assistants were recruited from two German speaking universities and born between 1985 and 1993.³ Research assistants were carefully trained and received detailed manuals on how to carry out the drop-offs. After walking into a building, research assistants were instructed to approach an employee at the counter and say: “*Hi, I found this [showing the wallet] on the street just around the corner.*” Then, they put the wallet on the counter and pushed it over to the employee: “*Somebody must have lost it. I’m in a hurry and have to go. Can you please*

3. In the Robustness Checks section on page 179 we assess the influence of research assistants and find no evidence that differences between experimenters are driving our main results.

*take care of it?*⁴ The research assistant subsequently left the building without leaving their contact details or asking for a receipt.⁵ This interaction was designed to minimize recipients' concerns about being punished, since there was no written proof that a wallet had been turned in. Furthermore, by telling recipients that the wallet was found outside the building around the corner, we avoided possible concerns that the owner might come back and look for the wallet in that exact location.

G. *Experimental Conditions*

Our primary experimental manipulation varied the amount of money in the wallet. In the “NoMoney” condition, the wallets only contained business cards, a shopping list, and a key. In the “Money” condition, the wallets also contained the equivalent of US \$13.45. We used local currencies, and to ensure comparability across countries we adjusted the amounts for purchasing power parity using data from the International Monetary Fund. Table A2.1 provides the exact amounts of money used in each country.

In three countries (the United Kingdom, Poland, and the United States), we conducted two additional treatment conditions. We ran a “BigMoney” condition that was identical to the Money condition but with the equivalent of US \$94.15 in the wallets (i.e., seven times the amount found in the Money condition). We also ran a “Money-NoKey” condition identical to the Money condition but the wallets did not contain a key. Because the key is only valuable to the owner, the Money-NoKey condition only varies the harm caused to the owner relative to the Money condition. This treatment therefore allows us to isolate the role of altruism in people's decision to return a lost wallet.

We randomly assigned treatments and owner identities to drop-off locations. Tables A2.2-A2.5 provide descriptive statistics and demonstrate that individual characteristics and situational factors are well balanced across treatments.

H. *Measuring Civic Honesty*

Our key outcome measure was whether a recipient contacted the owner to return the wallet. We created our own email server to collect responses. The business cards in each wallet had a unique email address that allowed us to automatically assign incoming emails to its respective drop-off location and to automatically send a reply message in the local language.

4. Recipients were always approached in English, but research assistants also used a translator app on their cell phones in case a recipient was not conversant in English.

5. Recipients rarely refused to take the wallet. The median rejection rate was less than 0.4%, with only five countries exhibiting rejection rates above 1% (and none greater than 5%). Columns (1) and (2) in Table A2.6 shows that rejection rates did not significantly differ between the Money, NoMoney, and Money-NoKey conditions. We find a marginally significant difference ($t_{2884} = 1.77$, $P = 0.077$) between the Money and the BigMoney condition, as shown in Column (2). Using χ^2 -tests, we find that only 3.3% of all possible pairwise country comparisons are significant at the 5% level after controlling for the false discovery rate (Benjamini and Hochberg, 1995).

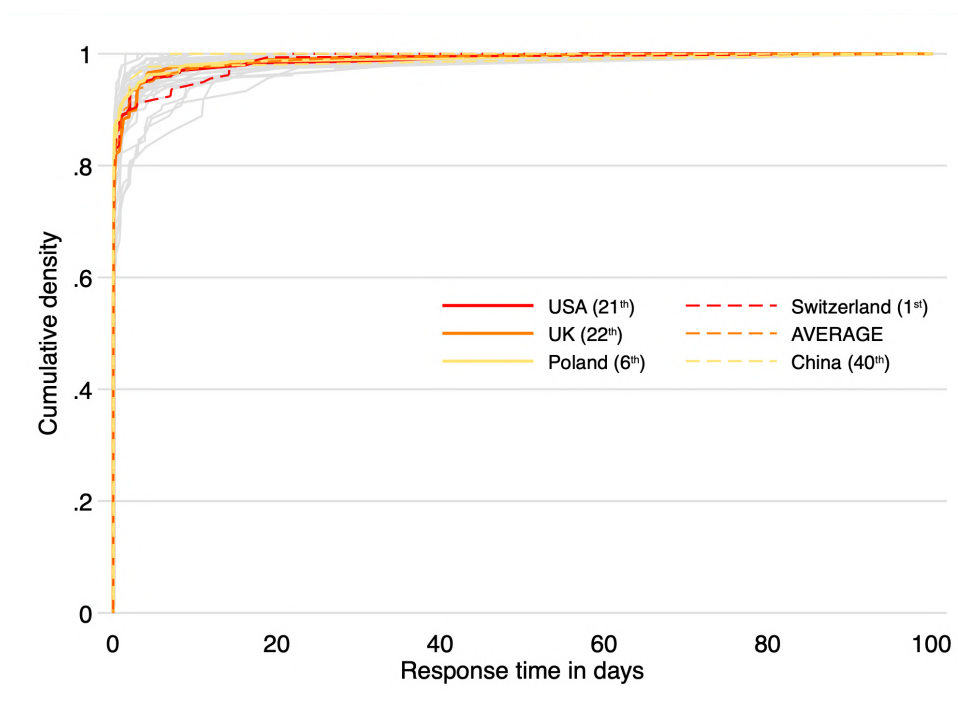


FIGURE A2.2

Cumulative Distribution Function of Response Times

Notes. Cumulative distribution function for the time elapsed between the drop-off of the wallets and the email responses from the recipients by country. The three main countries, and the countries with the highest and lowest response rate in the NoMoney condition are highlighted (ranking in parentheses).

The following reply message was sent three hours and fifteen minutes after receiving an email from the recipients: *“Hello, thank you very much for your email. I really appreciate your help. Unfortunately, I have already left town. The content of the business card holder and the key are not important to me. You can keep all of it or donate it to charity. Best regards, [firstname]/[lastname].”* If present, we specifically mentioned the key because recipients would frequently inquire about the key in follow-up emails. If multiple emails were sent to the same email address then we flagged them for review by a research assistant. The majority of these emails did not necessitate further action.

Besides automating part of the data collection, the use of a private email server allowed us to register attempts to return a wallet even if the email address was spelled incorrectly. As long as the domain name was spelled correctly, a research assistant could manually reassign the email to the correct drop-off location. Common errors involve forgetting the dot in the email address or using a similar name, such as “lars-andersen” or “lars.andresen” instead of “lars.andersen.”

We recorded emails that were sent within 100 days after the drop-offs. The median response time was roughly 26 minutes across all countries, and about 88% of emails arrived within 24 hours (see Figure A2.2). Table A2.7 shows that response times did not significantly differ across

treatments. Moreover, we find little variation in response times between countries. Using two-sample t -tests, we find that only 1.5% of all possible pairwise country comparisons are significant at the 5% level (FDR-adjusted P -values). Average response times and response rates by country were not significantly correlated (Spearman's $\rho = 0.162$, $P = 0.319$).

I. Measuring Recipient Characteristics and Situational Factors

Upon leaving the locations, our research assistants filled out a short survey to collect additional information about the drop-offs. This data allowed to account for incidental factors that varied across locations. Research assistants recorded the following information:

- *Recipient gender.* Research assistants took note of the recipient's gender, which was coded as 0 for female and 1 for male.
- *Recipient age.* Research assistants estimated the recipient's approximate age on a 6-point scale: < 20 , 20-30, 30-40, 40-50, 50-60, > 60 . For all analyses using age we use a median split dummy variable in which we coded as 1 if the recipient was estimated to be 40 years or older and 0 otherwise. We used a median split for purposes of simplicity; using a set of indicator variables for each age category does not meaningfully affect any of the treatment effects we report.
- *Busyness.* Research assistants estimated how busy the recipient was on a 7-point scale from "not at all" (0) to "very busy" (6).
- *Local recipient.* Research assistants assessed whether the recipient was a foreigner on a 7-point scale from "local" (0) to "unclear" (3) to "foreigner" (6). We coded this variable as 1 if the recipient was rated below the midpoint of the scale and 0 otherwise. We used this indicator variable for purposes of simplicity; treating this local residency as a continuous variable does not meaningfully affect any of the treatment effects we report.
- *No English.* Whether the research assistant had to use a different language than English to communicate with the recipient (using a mobile phone app). We coded this as 0 if the recipient understood English, and 1 if the recipient did not. This variable was always coded as 0 if a region is English speaking.
- *Recipient understood situation.* Research assistants assessed the extent the recipient understood the situation on a 7-point scale from "not at all" (0) to "fully understood" (6). We only collected this information after finishing data collection in Poland and the UK.
- *Friendliness.* Research assistants assessed the friendliness of the recipient on a 7-point scale from "very unfriendly" (0) to "very friendly" (6). We only collected this information after finishing data collection in Croatia, Greece, Poland, Romania, Serbia, and the UK.

- *Computer*. Research assistants noted if there was a computer at the recipient's desk (0 = computer absent, 1 = computer present).
- *Coworkers*. Research assistants took note of how many employees participated in or closely witnessed the exchange. They had the following response options: one, two, three, or more than three. For all analyses using this variable we coded this variable as 1 if multiple coworkers participated or closely witnessed the exchange and 0 otherwise. We used this indicator variable for purposes of simplicity; using a set of indicator variables for each response option does not meaningfully affect any of the treatment effects we report.
- *Other bystanders*. Research assistants took note of how many other people could witness the drop-off. They had the following response options: none, fewer than five, five or more. For all analyses using presence of bystanders we coded this variable as 1 if any bystanders were present and 0 otherwise. We used this indicator variable for purposes of simplicity; using a set of indicator variables for each response option does not meaningfully affect any of the treatment effects we report.
- *Security camera*. Research assistants took note of whether a security camera was visible in the room (0 = no camera visible, 1 = camera visible). We only collected this information after finishing data collection in Poland and the UK.
- *Security guard*. Research assistants took note of whether a security was present (0 = guard present, 1 = guard absent). We only collected this information after finishing data collection in Croatia, Greece, Poland, Romania, Serbia, and the UK.

J. Country-Level Correlates of Civic Honesty

As a supplement to our experimental study, we also examined country-level predictors of civic honesty. We examined how rates of civic honesty vary according to the following set of country-level characteristics:

- *Country GDP*. Logarithm of country gross domestic product based on purchasing-power-parity per capita in 2010 from the IMF World Economic Outlook (Fund, 2015).
- *Log. soil fertility*. Logarithm of soil suitability, obtained from Ashraf and Galor (2013). The data is originally from Ramankutty et al. (2002) who estimated soil suitability at half-degree resolution based on soil pH and soil carbon density. The data was then aggregated at the country-level by Michalopoulos (2012). Missing data on Serbia has been replaced with data from Yugoslavia.
- *Log. abs. latitude*. Logarithm of the absolute latitude of a country's approximate geodesic centroid, obtained from Ashraf and Galor (2013). The data is originally from the CIA's *World Factbook*. Missing data on Serbia has been replaced with data from Yugoslavia.

- *Distance to waterway.* Distance (in 100 km) to the nearest ice-free coastline or sea-navigable river, obtained from Ashraf and Galor (2013). The data is originally from Gallup et al. (1999). Missing data on Serbia has been replaced with data from Yugoslavia.
- *Temperature.* Average monthly temperature (in Celsius degrees) of a country between 1961 and 1990, obtained from Ashraf and Galor (2013). The data is originally from the G-ECON project (Nordhaus, 2006). Missing data on Serbia has been replaced with data from Yugoslavia.
- *Precipitation.* Average monthly precipitation (in mm per month) of a country between 1961 and 1990, obtained from Ashraf and Galor (2013). The data is originally from the G-ECON project (Nordhaus, 2006). Missing data on Serbia has been replaced with data from Yugoslavia.
- *Mean elevation.* Mean elevation of a country (in km) above sea level, obtained from Ashraf and Galor (2013). The data is originally from the G-ECON project (Nordhaus, 2006). Missing data on Serbia has been replaced with data from Yugoslavia.
- *Terrain roughness.* Degree of terrain roughness, obtained from Ashraf and Galor (2013). The data is originally from the G-ECON project (Nordhaus, 2006). Missing data on Serbia has been replaced with data from Yugoslavia.
- *Temperature (Volatility).* Ancestry adjusted volatility of temperature between 1900 and 2000. Based on the *Climatic Research Unit (CRU)* database, and constructed using the method outlined in Durante (2010). The variable is obtained from Galor and Özak (2016). Missing data on Serbia has been replaced with data from Yugoslavia.
- *Precipitation (Volatility).* Ancestry adjusted volatility of precipitation between 1900 and 2000. Based on the *Climatic Research Unit (CRU)* database, and constructed using the method outlined in Durante (2010). The variable is obtained from Galor and Özak (2016). Missing data on Serbia has been replaced with data from Yugoslavia.
- *Pathogen prevalence.* Historic prevalence of nine infectious diseases (leishmaniasis, schistosomes, trypanosomes, leprosy, malaria, typhus, filariae, dengue, and tuberculosis) based on epidemiological atlases from the first half of the 20th century, as constructed by Murray and Schaller (2010). The variable is obtained from Gorodnichenko and Roland (2017). Data on Kazakhstan (and several other countries not covered by our study) has been fitted based on an index of seven pathogens (excluding leprosy and tuberculosis) (Gorodnichenko and Roland, 2017).
- *Pronoun drop not allowed.* Share of individuals that speak a language that does not allow dropping the first-person pronoun (i.e., "I"), thereby putting more emphasis on the

individual (Tabellini, 2008a). The variable was obtained from Tabellini (2008a). The data is originally from Kashima and Kashima (1998). Data on Croatia, Kazakhstan, Morocco, and Serbia has been manually completed based on the major languages using data from the *World Atlas of Language Structures (WALS)*.

- *Politeness distinction.* Share of individuals that speak a language that prescribes the use of different pronouns (e.g., “tu” and “vous” in French) depending on the relationship between the speakers. This is a trait that has been linked to hierarchy and power distance (Tabellini, 2008a). The variable was obtained from Tabellini (2008a). The data is originally from Kashima and Kashima (1998). Data on Croatia, Kazakhstan, Morocco, and Serbia has been manually completed based on the major languages using data from the *World Atlas of Language Structures (WALS)*.
- *Weak future time reference.* Share of individuals that speak a language with a weak future time reference, obtained from Chen (2013). Languages with a weak future time reference allow the speaker to use the same grammatical tense to speak about present and future events instead of having a grammatically distinct future tense. Data on Brazil, Morocco, Peru, Serbia, South Africa, Indonesia, Ghana, Kenya, Kazakhstan, India, and the United Arab Emirates have been manually completed based on the major languages using data from the *World Atlas of Language Structures (WALS)*.
- *Share of protestants.* Percentage of a country’s population that is protestant, obtained from Ashraf and Galor (2013). The data is originally from La Porta et al. (1999). Missing data on Serbia has been manually completed using data from the Serbian census in 2002.
- *Family ties.* Strength of family ties calculated following Alesina and Giuliano (2010). The variable is the first principal component of three family-related questions in the *World Value Survey (WVS)*: (i) “For each of the following, indicate how important it is in your life. - Family:”, on a 4-point scale from 1 (*not important at all*) to 4 (*very important*), (ii) “With which of these two statements do you tend to agree? A: One does not have the duty to respect and love parents who have not earned it; B: Regardless of what the qualities and faults of one’s parents are, one must always love and respect them.” (iii) “Which of the following statements best describes your views about parents’ responsibilities to their children? A: Parents have a life of their own and should not be asked to sacrifice their own well-being for the sake of their children; B: It is the parents’ duty to do their best for their children even at the expense of their own well being.”
- *State history.* State history index (Bockstette et al., 2002). For each period of 50 years from year 1 C.E. to 1950, a country’s experience with supra-tribal government is coded for (i) the existence of a government above the tribal level, (ii) whether said government was foreign or locally based, and (iii) how much of the current country it ruled. A discount

factor of 5% for each 50 years is applied so that more recent experience with statehood are weighted more heavily in the index. The variable is obtained from Spolaore and Wacziarg (2013).

- *Years of democracy.* Years since the polity score in the *Polity IV* data set is strictly above zero, starting from 1800 or the year of independence for countries that became independent later. The polity score is defined by subtracting the autocracy score from the democracy score and ranges from “strongly democratic” (10) to “strongly autocratic” (−10).
- *Executive constraints.* Constraints on executive scale from the *Polity IV* data set. The scale takes values from “unlimited authority” (1) to “executive parity or subordination” (7), the later being defined as a situation in which “accountability groups have effective authority equal to or greater than the executive in most areas of activity.”
- *Judicial independence.* Judicial independence as of 1995, obtained from Glaeser et al. (2004). The data is originally from La Porta et al. (2004) who defined the variable as the sum of three sub-scales measuring (i) tenure of supreme court judges, (ii) tenure of the highest ranked judges ruling on administrative cases, and (iii) the existence of case law. The variable is normalized to range from zero to one.
- *Constitutional review.* Constitutional review as of 1995, obtained from Glaeser et al. (2004). The data is originally from La Porta et al. (2004) who defined the variable as the sum of two sub-scales measuring (i) the extent to which judges of the supreme or constitutional court can review the constitutionality of laws and (ii) how difficult it is to change the constitution. The variable is normalized to range from zero to one.
- *Electoral rule: Plurality.* Percentage of years between 1975 and 2000 in which a first-past-the-post or winner-takes-all system was used to elect legislators, obtained from Glaeser et al. (2004). The data is originally from Beck et al. (2001).
- *Electoral rule: Proportionality.* Percentage of years between 1975 and 2000 in which a proportional system was used to elect legislators, i.e., legislators were elected based on the share of votes that their party received in an election. The variables is obtained from Glaeser et al. (2004). The data is originally from Beck et al. (2001).
- *Primary education 1920.* Primary school enrollment in 1920, obtained from Benavot and Riddle (1988).

TABLE A2.1

Country	Cities	Treatments	Names	Shopping List	Languages	Email	N
Argentina (16 Jul. 2015 – 8 Aug. 2015)	Buenos Aires	Money (ARS 48.50)	Eduardo Martinez	dulce de leche	Spanish		400
	Córdoba	NoMoney	Guillermo Garcia	pan			
	Mar del Plata		Ignacio Lombardi	costilla			
	Mendoza			lata duraznos			
	Rosario						
Australia (4 May 2015 – 25 May 2015)	Salta						
	San Miguel de Tucumán						
	Santa Fe						
	Adelaide	Money (AUD 20)	Jack Williams	milk	English		399 ^a
	Brisbane	NoMoney	James Smith	bread			
Brazil (25 Jul. 2015 – 13 Aug. 2015)	Canberra		William Jones	noodles			
	Goldcoast			bananas			
	Melbourne						
	Newcastle						
	Perth						
Sydney							
	Belo Horizonte	Money (BRL 21.75)	Gabriel Pereira Lima	leite	Portuguese		399 ^a
	Brasília	NoMoney	Lucas de Oliveira Souza	pão			
	Curitiba		Rafael da Luz Santos	macarrão			
	Fortaleza			frutas			
Canada (10 Sep. 2015 – 6 Oct. 2015)	Manaus						
	Rio de Janeiro						
	Salvador						
	São Paulo						
	Calgary	Money (CAD 16.50)	David Smith	milk	English		400
Edmonton	NoMoney	Jacob Brown	bread				
	Ottawa		Robert Wilson	pasta			
	Toronto			bananas			
	Vancouver						
	Montréal		Alexandre Gagnon	lait	French		
Québec		Olivier Roy	pain				
		Thomas Tremblay	pâtes				
				bananes			
Continued							

Continued

TABLE A2.1

Country	Cities	Treatments	Names	Shopping List	Languages	Email	N
Chile (16 Nov. 2016 – 13 Dec. 2016)	Antofagasta	Money (CLP 4650) NoMoney	Diego Rojas Francisco Muñoz Jorge Gonzalez	leche pan chocolo plátanos	Spanish		396 ^b
	Concepción						
	La Serena						
	Rancagua						
	Santiago de Chile						
	Talca						
	Temuco						
	Valparaíso						
China (7 Jul. 2015 – 27 Jul. 2015)	Beijing	Money (RMB 49) NoMoney	Chang Wei Li Qiang Wang Lei	矿泉水 包子 方便面 苹果	Chinese		400
	Chengdu						
	Guangzhou						
	Hangzhou						
	Shanghai						
	Shenzhen						
	Tianjin						
	Xi'an						
Croatia (25 Mar. 2014 – 4 Apr. 2014)	Osijek	Money (HRK 53.50) NoMoney	Ivan Kovačević Marko Horvat Tomislav Babić	mlijeka kruha tjestenine banane	Croatian		300
	Rijeka						
	Split						
	Zadar						
	Zagreb						
Czech Republic (21 Oct. 2014 – 7 Nov. 2014)	Brno	Money (CZK 170) NoMoney	Marek Pospíšil Václav Korbela — ^c	mléka chleba těstovin banány	Czech		400
	Liberec						
	Olomouc						
	Ostrava						
	Plzeň						
	Praha						
Denmark (15 Jul. 2014 – 25 Jul. 2014)	Aalborg	Money (DKK 109) NoMoney	Christian Rasmussen Lars Andersen Søren Jensen	mælk brød pasta bananen	Danish		300
	Århus						
	København						
	Odense						
Continued							

TABLE A2.1

Continued

TABLE A2.1

Continued

TABLE A2.1

Country	Cities	Treatments	Names	Shopping List	Languages	Email	N
Italy (15 Jun. 2014 – 11 Jul. 2014)	Bari	Money (EUR 10.75) NoMoney	Antonio Gallo Giuseppe Russo Roberto Bianchi	latte pane pasta banane	Italian		400
	Bologna						
	Catania						
	Firenze						
	Genova						
	Messina						
	Milano						
	Napoli						
	Padova						
	Palermo						
	Roma						
	Taranto						
	Torino						
	Trieste						
	Venezia						
	Verona						
Kazakhstan (4 Jun. 2016 – 16 Jun. 2016)	Almaty	Money (KZT 1203) NoMoney	Alexey Omarov Bekzat Akhmetov Kirill Osranov	сүт нан кеспе банан	Kazakh		400
	Astana						
	Karaganda						
	Oskemen						
	Pavlodar						
	Semey						
	Shymkent						
	Taraz						
Kenya (10 Nov. 2015 – 24 Nov. 2015)	Nairobi	Money (KES 507) NoMoney	John Omondi Peter Kihiga Samuel Jabali	iria ngima chapati mũraru chak kuon chapati rabolo	English		274 ^e
	Nakuru						
	Nyeri						
	Eldoret						
	Kisumu						
Continued							

TABLE A2.1

Country	Cities	Treatments	Names	Shopping List	Languages	Email	N
Malaysia (23 May 2016 – 13 Jun. 2016)	Mombasa _e			maziwa ugali chapati ndizi			
	George Town	Money (MYR 10.15)	Amir bin Roslan		Malay		400
	Ipoh	NoMoney	Farid bin Azhari				
	Johor Bahru		Ramlee bin Khadir				
	Kota Kinabalu						
Mexico (7 Sep. 2015 – 28 Sep. 2015)	Kuala Lumpur						
	Melaka						
	Petaling Jaya						
	Shah Alam						
	Chihuahua	Money (MXN 105)	Carlos García	leche	Spanish		400
Morocco (25 May 2015 – 12 Jun. 2015)	Guadalajara	NoMoney	Daniel Martínez	pan			
	León		José Hernandez	pasta			
	Mérida			plátanos			
	Mexico City						
	Monterrey						
Morocco (25 May 2015 – 12 Jun. 2015)	Puebla						
	Tijuana						
	Agadir	Money (MAD 49)	Ahmed el Mernissi	حليب	French		402 ^a
	Casablanca	NoMoney	Mohamed Bennani	خبز			
	Fez		Yassine ben Aissa	الكسكس موز			
	Quneitra						
	Marrakesh						
	Meknès						
	Rabat						
	Tangier						
	Tétouan						
Continued							

TABLE A2.1
SAMPLE OVERVIEW (CONTINUED)

TABLE A2.1

Country	Cities	Treatments	Names	Shopping List	Languages	Email	N
Poland (25 Jul. 2013 – 29 Aug. 2013)	Białystok	BigMoney (PLN 175) Money (PLN 25) Money-NoKey (PLN 25) NoMoney	Edward Kowalski	mleka	Polish		800
	Bydgoszcz		Marek Nowak	chleb			
	Bytom		Paweł Wiśniewski	makaron			
	Częstochowa			banany			
	Gdańsk						
	Gdynia						
	Gliwice						
	Katowice						
	Kielce						
	Krakow						
	Łódź						
	Lublin						
	Opole						
	Poznan						
Radom							
Sosnowiec							
Szczecin							
Toruń							
Warszawa							
Wroclaw							
Portugal (15 May 2014 – 31 May 2014)	Braga	Money (EUR 8.50) NoMoney	João Silva Fernandes	leite	Portuguese		400
	Coimbra		Miguel Ferreira Rodrigues	pão			
	Faro		Rodrigo Santos Pereira	pasta			
	Lisboa			bananas			
	Porto						
	Setúbal						
Romania (25 Mar. 2014 – 15 Apr. 2014)	Braşov	Money (RON 28) NoMoney	Andrei Popescu	lapte	Romanian		400
	Bucureşti		Constantin Radu	pâine			
	Cluj-Napoca		Gheorghe Matei	paste			
	Constanţa			banane			
	Galaţi						
	Iaşi						
Timișoara							
Continued							

TABLE A2.1

Country	Cities	Treatments	Names	Shopping List	Languages	Email	N
Russia (4 Aug. 2015 – 18 Aug. 2015)	Kazan	Money (RUB 258)	Daniil Smirnov	молоко	Russian		302 ^a
	Moscow	NoMoney	Dimitri Ivanov	хлеб			
	Nizhny Novgorod		Ivan Kuznetsov	лапша			
	Novosibirsk			бананы			
	Omsk						
Serbia (7 Apr. 2014 – 17 Apr. 2014)	Belgrade	Money (DIN 612)	Dragan Pavlović	млека	Serbian		300
	Kragujevac	NoMoney	Nikola Stojanović	хлеб			
	Niš		Vladimir Nikolic	тестенина			
	Novi Sad			банане			
	Subotica						
South Africa (13 Jan. 2016 – 11 Feb. 2016)	Bloemfontein	Money (ZAR 69)	Johan Fourie	milk	English		399 ^a
	Cape Town	NoMoney	Michael Botha	bread			
	Durban		Thabo Molefe	rice			
	East London		Tshepo Mokwena	bananas			
	Johannesburg						
Spain (13 May 2014 – 25 Jun. 2014)	Pietermaritzburg						
	Port Elizabeth						
	Pretoria						
	A Coruña	Money (EUR 9.50)	Antonio García González	leche	Spanish		400
	Alicante	NoMoney	José Fernández García	pan			
	Barcelona		Manuel González Fernández	pasta			
	Bilbao			plátanos			
	Córdoba						
	Gijón						
	Madrid						
	Málaga						
	Murcia						
	Palma						
	Sevilla						
	Valencia						
Valladolid							
Vigo							
Zaragoza							

Continued

TABLE A2.1

Country	Cities	Treatments	Names	Shopping List	Languages	Email	N					
Sweden (21 Aug. 2014 – 5 Sep. 2014)	Göteborg	Money (SEK 115) NoMoney	Anders Johansson	mjök	Swedish		400					
	Helsingborg		Lars Andersson	bröd								
	Jönköping		Mikael Karlsson	pasta								
	Linköping			bananer								
	Lund											
	Malmö											
	Norrköping											
Switzerland (8 Sep. 2014 – 26 Sep. 2014)	Stockholm	Money (CHF 20.75) NoMoney	Daniel Martin	Milch	German		399 ^a					
	Uppsala		Marco Schwarz	Brot								
	Basel		– c	Pasta								
	Bern			Bananen								
	Luzern											
	St. Gallen											
	Winterthur											
	Zürich											
	Geneva			lait				French				
	Lausanne			pain								
Thailand (16 May 2015 – 12 Jun. 2015)	Bangkok	Money (THB 166) NoMoney	Charoen Bongkot	พริกแกง	Thai		400					
	Chiang Mai		Somsak Banyat	ซอสถั่วเหลือง								
	Hat Yai		Thongchai Malechan	ข้าว								
	Khon Kaen			มะม่วง								
	Nakhon Ratchasima											
	Nakhon Si Thammarat											
	Ubon Ratchathani											
	Udon Thani											
	Turkey (26 Jun. 2014 – 11 Jul. 2014)		Adana	Money (TRY 16) NoMoney				Hakan Kaya	süt	Turkish		400
			Ankara					Mesut Demir	ekmek			
Antalya		Mustafa Şahin	makarna									
Gaziantep			muz									
İstanbul												
İzmir												
Konya												

Continued

TABLE A2.1
SAMPLE OVERVIEW (CONTINUED)

Country	Cities	Treatments	Names	Shopping List	Languages	Email	N
US (17 Aug. 2015 – 17 Oct. 2015)	Albuquerque	BigMoney (USD 94.15)	Brad O'Brien	milk	English		1000
	Boston	Money (USD 13.45)	Brett Miller	bread			
	Charlotte	Money-NoKey (USD 13.45)	Connor Baker	pasta			
	Chicago	NoMoney		bananas			
	Columbus						
	Denver						
	Houston						
	Indianapolis						
	Jacksonville						
	Las Vegas						
	Los Angeles						
	Louisville						
	Memphis						
	Milwaukee						
	Nashville						
	New York						
	Oklahoma City						
	Philadelphia						
	Phoenix						
	Portland						
40 Countries	San Antonio						17303
	San Diego						
	Seattle						
	Tucson						
	Washington, D.C						

Notes.

^a There are minor deviations (+/- 2) from the target sample size in some countries due to rounding errors in the planning of the sample size for each city or because no suitable replacement for a closed drop-off location could be found in time.

^b A handling mistake made it impossible to identify the drop-off locations of four wallets.

^c We used only two instead of three names in the Czech Republic and Switzerland. We used real names in these countries to be able to pick up the wallets as part of the internal validation.

^d Coimbatore was chosen as a last minute replacement for Chennai after the February 2015 flooding.

^e We exclude the last city that we visited (Malindi) because the experimenter was arrested and interrogated by the military police for suspicious activity.

^f We excluded several observations because of a printing error concerning the email address on the business cards.

TABLE A2.2
DESCRIPTIVE STATISTICS AND RANDOMIZATION CHECK FOR THE UNITED KINGDOM

	NoMoney		Money		Money-NoKey		BigMoney		Total sample		<i>P</i> -value
	mean	<i>SD</i>	mean	<i>SD</i>	mean	<i>SD</i>	mean	<i>SD</i>	mean	<i>SD</i>	
Age 40+	0.335	(0.473)	0.326	(0.469)	0.355	(0.480)	0.382	(0.487)	0.343	(0.475)	0.538
Male	0.385	(0.488)	0.381	(0.486)	0.360	(0.481)	0.357	(0.480)	0.374	(0.484)	0.890
Computer	0.865	(0.343)	0.805	(0.397)	0.810	(0.393)	0.819	(0.386)	0.819	(0.385)	0.298
Coworkers	0.195	(0.397)	0.189	(0.392)	0.195	(0.397)	0.211	(0.409)	0.195	(0.397)	0.934
Other bystanders	0.725	(0.448)	0.775	(0.418)	0.795	(0.405)	0.809	(0.394)	0.776	(0.417)	0.199
Hotel	0.230	(0.422)	0.236	(0.425)	0.250	(0.434)	0.261	(0.440)	0.242	(0.429)	0.868
Bank	0.250	(0.434)	0.248	(0.432)	0.255	(0.437)	0.221	(0.416)	0.245	(0.430)	0.857
Cultural	0.225	(0.419)	0.242	(0.429)	0.225	(0.419)	0.216	(0.413)	0.231	(0.422)	0.875
Public	0.190	(0.393)	0.178	(0.383)	0.175	(0.381)	0.196	(0.398)	0.183	(0.387)	0.928
Postal	0.105	(0.307)	0.096	(0.294)	0.095	(0.294)	0.106	(0.308)	0.099	(0.299)	0.964
Local recipient	0.905	(0.294)	0.927	(0.261)	0.915	(0.280)	0.930	(0.256)	0.921	(0.269)	0.739
No English	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	1.000
Busy	2.145	(1.769)	2.201	(1.900)	2.180	(1.709)	2.166	(1.844)	2.181	(1.832)	0.993
Observations	200		533		200		199		1,132		

Notes. We coded “Age 40+” as 1 if the recipient was judged to be 40 years or older, and 0 otherwise. “Male” was coded as 1 if the recipient was male and 0 otherwise. “Computer” was coded as 1 if there was a computer present at the recipient’s desk, and 0 otherwise. “Coworkers” was coded as 1 if a recipient’s coworkers participated in the exchange, and 0 otherwise. “Other bystanders” was coded as 1 if any bystanders witness the exchange and 0 otherwise. The variables “Hotel,” “Bank,” “Cultural,” “Public,” and “Postal” represent the five types of institutions in which the experiments were performed. “Local recipient” was coded as 1 if the recipient was judged to be a local resident, and 0 otherwise. “No English” was coded as 1 if the recipient was not able to communicate with the research assistant in English, and 0 otherwise. “Busy” was a rating of how busy the recipient was when the wallet was turned in, on a scale from 0 (*not at all*) to 6 (*very busy*). The last column presents *P*-values for the null hypothesis of perfect randomization (χ^2 -tests, except for “busy” where we perform a Kruskal-Wallis *H* test).

TABLE A2.3
DESCRIPTIVE STATISTICS AND RANDOMIZATION CHECK FOR POLAND

	NoMoney		Money		Money-NoKey		BigMoney		Total sample		P-value
	mean	SD	mean	SD	mean	SD	mean	SD	mean	SD	
Age 40+	0.263 ⁽²⁾	(0.441)	0.286 ⁽¹⁾	(0.453)	0.275	(0.448)	0.350 ⁽³⁾	(0.478)	0.293 ⁽⁶⁾	(0.456)	0.226
Male	0.288 ⁽²⁾	(0.454)	0.352 ⁽¹⁾	(0.479)	0.265	(0.442)	0.313 ⁽²⁾	(0.465)	0.304 ⁽⁵⁾	(0.460)	0.272
Computer	0.894 ⁽²⁾	(0.309)	0.839 ⁽¹⁾	(0.368)	0.905	(0.294)	0.864 ⁽²⁾	(0.344)	0.875 ⁽⁵⁾	(0.330)	0.181
Coworkers	0.268 ⁽²⁾	(0.444)	0.286 ⁽¹⁾	(0.453)	0.360	(0.481)	0.333 ⁽²⁾	(0.473)	0.312 ⁽⁵⁾	(0.464)	0.173
Other bystanders	0.763 ⁽²⁾	(0.427)	0.673 ⁽¹⁾	(0.470)	0.665	(0.473)	0.732 ⁽²⁾	(0.444)	0.708 ⁽⁵⁾	(0.455)	0.095
Hotel	0.175	(0.381)	0.155	(0.363)	0.155	(0.363)	0.120	(0.326)	0.151	(0.359)	0.485
Bank	0.340	(0.475)	0.370	(0.484)	0.375	(0.485)	0.345	(0.477)	0.357	(0.480)	0.848
Cultural	0.165	(0.372)	0.145	(0.353)	0.150	(0.358)	0.175	(0.381)	0.159	(0.366)	0.837
Public	0.190	(0.393)	0.210	(0.408)	0.200	(0.401)	0.230	(0.422)	0.207	(0.406)	0.786
Postal	0.130	(0.337)	0.120	(0.326)	0.120	(0.326)	0.130	(0.337)	0.125	(0.331)	0.980
Local recipient	0.895	(0.307)	0.930	(0.256)	0.950	(0.218)	0.900	(0.301)	0.919	(0.273)	0.144
No English	0.515	(0.501)	0.625	(0.485)	0.560	(0.498)	0.605	(0.490)	0.576	(0.494)	0.116
Busy	2.827 ⁽³⁾	(1.964)	2.392 ⁽¹⁾	(1.906)	2.407 ⁽¹⁾	(1.912)	2.601 ⁽²⁾	(1.961)	2.556 ⁽⁷⁾	(1.940)	0.096
Observations	200		200		200		200		800		

Notes. We coded “Age 40+” as 1 if the recipient was judged to be 40 years or older, and 0 otherwise. “Male” was coded as 1 if the recipient was male and 0 otherwise. “Computer” was coded as 1 if there was a computer present at the recipient’s desk, and 0 otherwise. “Coworkers” was coded as 1 if a recipient’s coworkers participated in the exchange, and 0 otherwise. “Other bystanders” was coded as 1 if any bystanders witness the exchange and 0 otherwise. The variables “Hotel,” “Bank,” “Cultural,” “Public,” and “Postal” represent the five types of institutions in which the experiments were performed. “Local recipient” was coded as 1 if the recipient was judged to be a local resident, and 0 otherwise. “No English” was coded as 1 if the recipient was not able to communicate with the research assistant in English, and 0 otherwise. “Busy” was a rating of how busy the recipient was when the wallet was turned in, on a scale from 0 (*not at all*) to 6 (*very busy*). The last column presents *P*-values for the null hypothesis of perfect randomization (χ^2 -tests, except for “busy” where we perform a Kruskal-Wallis *H* test). There are a few missing answers in the drop-off survey. Numbers in superscript parenthesis indicate the number of drop-offs that are missing in each condition.

TABLE A2.4
DESCRIPTIVE STATISTICS AND RANDOMIZATION CHECK FOR THE UNITED STATES

	NoMoney		Money		Money-NoKey		BigMoney		Total sample		<i>P</i> -value
	mean	<i>SD</i>	mean	<i>SD</i>	mean	<i>SD</i>	mean	<i>SD</i>	mean	<i>SD</i>	
Age 40+	0.487	(0.501)	0.453	(0.499)	0.465	(0.500)	0.450	(0.499)	0.465	(0.499)	0.823
Male	0.377	(0.485)	0.417	(0.494)	0.385	(0.488)	0.535	(0.500)	0.422	(0.494)	0.003
Computer	0.860	(0.348)	0.890	(0.313)	0.910	(0.287)	0.865	(0.343)	0.880	(0.325)	0.314
Coworkers	0.220	(0.415)	0.283	(0.451)	0.290	(0.455)	0.275	(0.448)	0.264	(0.441)	0.223
Other bystanders	0.710	(0.455)	0.720	(0.450)	0.700	(0.459)	0.725	(0.448)	0.714	(0.452)	0.943
Hotel	0.230	(0.422)	0.207	(0.406)	0.270	(0.445)	0.230	(0.422)	0.231	(0.422)	0.438
Bank	0.237	(0.426)	0.223	(0.417)	0.185	(0.389)	0.220	(0.415)	0.219	(0.414)	0.586
Cultural	0.247	(0.432)	0.250	(0.434)	0.220	(0.415)	0.210	(0.408)	0.235	(0.424)	0.671
Public	0.193	(0.396)	0.190	(0.393)	0.270	(0.445)	0.255	(0.437)	0.220	(0.414)	0.067
Postal	0.093	(0.291)	0.130	(0.337)	0.055	(0.229)	0.085	(0.280)	0.095	(0.293)	0.041
Security camera	0.783	(0.413)	0.830	(0.376)	0.835	(0.372)	0.835	(0.372)	0.818	(0.386)	0.322
Security guard	0.173	(0.379)	0.210	(0.408)	0.255	(0.437)	0.220	(0.415)	0.210	(0.408)	0.172
Local recipient	0.947	(0.225)	0.957	(0.204)	0.945	(0.229)	0.945	(0.229)	0.949	(0.220)	0.912
No English	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)	1.000
Understood situation	5.953	(0.438)	5.990	(0.173)	6.000	(0.000)	5.970	(0.316)	5.977	(0.294)	0.264
Busy	1.333	(1.661)	1.400	(1.638)	1.145	(1.545)	1.130	(1.447)	1.275	(1.592)	0.230
Observations	300		300		200		200		1,000		

Notes. We coded “Age 40+” as 1 if the recipient was judged to be 40 years or older, and 0 otherwise. “Male” was coded as 1 if the recipient was male and 0 otherwise. “Computer” was coded as 1 if there was a computer present at the recipient’s desk, and 0 otherwise. “Coworkers” was coded as 1 if a recipient’s coworkers participated in the exchange, and 0 otherwise. “Other bystanders” was coded as 1 if any bystanders witness the exchange and 0 otherwise. The variables “Hotel,” “Bank,” “Cultural,” “Public,” and “Postal” represent the five types of institutions in which the experiments were performed. “Security camera” was coded as 1 if a security camera was visible during the exchange, and 0 otherwise. “Security guard” was coded as 1 if a security guard was present, and 0 otherwise. “Local recipient” was coded as 1 if the recipient was judged to be a local resident, and 0 otherwise. “No English” was coded as 1 if the recipient was not able to communicate with the research assistant in English, and 0 otherwise. “Understood situation” was a rating of whether the recipient understood the situation on a scale from 0 (*not at all*) to 6 (*fully understood*). “Busy” was a rating of how busy the recipient was when the wallet was turned in, on a scale from 0 (*not at all*) to 6 (*very busy*). The last column presents *P*-values for the null hypothesis of perfect randomization (χ^2 -tests, except for “understood situation” and “busy” where we perform a Kruskal-Wallis *H* test).

TABLE A2.5
DESCRIPTIVE STATISTICS AND RANDOMIZATION CHECK FOR THE GLOBAL DATA

	NoMoney		Money		Total sample		<i>P</i> -value
	mean	<i>SD</i>	mean	<i>SD</i>	mean	<i>SD</i>	
Age 40+	0.408 ⁽²⁾	(0.491)	0.411 ⁽²⁾	(0.492)	0.409 ⁽⁴⁾	(0.492)	0.684
Male	0.464 ⁽²⁾	(0.499)	0.463 ⁽¹⁾	(0.499)	0.463 ⁽³⁾	(0.499)	0.967
Computer	0.773 ⁽²⁾	(0.419)	0.775 ⁽³⁾	(0.418)	0.774 ⁽⁵⁾	(0.418)	0.778
Coworkers	0.332 ⁽²⁾	(0.471)	0.343 ⁽¹⁾	(0.475)	0.338 ⁽³⁾	(0.473)	0.155
Other bystanders	0.645 ⁽²⁾	(0.479)	0.656 ⁽²⁾	(0.475)	0.651 ⁽⁴⁾	(0.477)	0.112
Hotel	0.218	(0.413)	0.217	(0.412)	0.218	(0.413)	0.855
Bank	0.233	(0.423)	0.231	(0.422)	0.232	(0.422)	0.806
Cultural	0.199	(0.400)	0.202	(0.401)	0.201	(0.401)	0.709
Public	0.210	(0.408)	0.211	(0.408)	0.211	(0.408)	0.926
Postal	0.139	(0.346)	0.139	(0.346)	0.139	(0.346)	0.981
Security camera	0.615 ⁽⁴⁰⁰⁾	(0.487)	0.619 ⁽⁷³³⁾	(0.486)	0.617 ^(1,133)	(0.486)	0.619
Security guard	0.255 ^(1,100)	(0.436)	0.262 ^(1,433)	(0.440)	0.259 ^(2,533)	(0.438)	0.324
Local recipient	0.936	(0.245)	0.932	(0.252)	0.934	(0.249)	0.349
No English	0.324	(0.468)	0.315	(0.464)	0.319	(0.466)	0.215
Understood situation	5.733 ⁽⁴²³⁾	(0.673)	5.748 ⁽⁷⁵⁷⁾	(0.631)	5.740 ^(1,180)	(0.652)	0.559
Busy	1.932 ⁽⁴⁾	(1.696)	1.961 ⁽²⁾	(1.715)	1.947 ⁽⁶⁾	(1.706)	0.333
Observations	7,890		8,214		16,104		

Notes. We coded "Age 40+" as 1 if the recipient was judged to be 40 years or older, and 0 otherwise. "Male" was coded as 1 if the recipient was male and 0 otherwise. "Computer" was coded as 1 if there was a computer present at the recipient's desk, and 0 otherwise. "Coworkers" was coded as 1 if a recipient's coworkers participated in the exchange, and 0 otherwise. "Other bystanders" was coded as 1 if any bystanders witness the exchange and 0 otherwise. The variables "Hotel," "Bank," "Cultural," "Public," and "Postal" represent the five types of institutions in which the experiments were performed. "Security camera" was coded as 1 if a security camera was visible during the exchange, and 0 otherwise. "Security guard" was coded as 1 if a security guard was present, and 0 otherwise. "Local recipient" was coded as 1 if the recipient was judged to be a local resident, and 0 otherwise. "No English" was coded as 1 if the recipient was not able to communicate with the research assistant in English, and 0 otherwise. "Understood situation" was a rating of whether the recipient understood the situation on a scale from 0 (*not at all*) to 6 (*fully understood*). "Busy" was a rating of how busy the recipient was when the wallet was turned in, on a scale from 0 (*not at all*) to 6 (*very busy*). The last column presents *P*-values for the null hypothesis of perfect randomization (χ^2 -tests, except for "understood situation" and "busy" where we perform a Kruskal-Wallis *H* test). Not all variables were collected in each country and there are a few missing answers in the drop-off survey. Numbers in superscript parenthesis indicate the number of drop-offs that are missing in each condition.

TABLE A2.6
ANALYSIS OF REJECTIONS

	All countries	UK, Poland, and US
	(1)	(2)
Money	0.182 (0.121)	-0.270 (0.380)
BigMoney		0.617 (0.559)
Money-NoKey		0.517 (0.547)
Constant	0.130 (0.161)	0.380 (0.427)
Controls:		
Institution FE	yes	yes
City FE	yes	yes
Money = BigMoney		0.077
Money = Money-NoKey		0.117
BigMoney = Money-NoKey		0.878
Wald test		0.184
Observations	16204	2959
Adjusted R^2	0.009	0.006

Notes. OLS estimates with robust standard errors in parentheses. Column (1) shows the results for treatment Money and NoMoney in all 40 countries. Column (2) shows the results for all four treatments in the United Kingdom, Poland, and the United States. The dependent variable is a dummy variable indicating whether the recipient refused to take the wallet. “Money,” “BigMoney,” and “Money-NoKey” are treatment indicators. The omitted category is the treatment “NoMoney.” All models include city and institution fixed effects. The bottom of the table reports P -values from t -tests for equality of the treatment coefficients and a Wald test of the joint significance of all treatments. Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A2.7
ANALYSIS OF RESPONSE TIMES

	All countries	UK, Poland, and US
	(1)	(2)
Money	−0.082 (0.131)	−0.053 (0.491)
BigMoney		−0.142 (0.497)
Money-NoKey		−0.232 (0.510)
Constant	0.760 (0.630)	−0.185 (0.492)
Controls:		
Institution FE	Yes	Yes
City FE	Yes	Yes
Money = BigMoney		0.840
Money = Money-NoKey		0.666
BigMoney = Money-NoKey		0.838
Wald test		0.964
Observations	7340	1711
Adjusted R^2	0.015	−0.004

Notes. OLS estimates with robust standard errors in parentheses. The dependent variable is the response time in days. Column (1) shows the results for treatment Money and NoMoney in all 40 countries. Column (2) shows the results for all four treatments in the United Kingdom, Poland, and the United States. The dependent variable is the response time in days. “Money,” “BigMoney,” and “Money-NoKey” are treatment indicators. The omitted category is the treatment “NoMoney.” All models include city and institution fixed effects. The bottom of the table reports P -values from t -tests for equality of the treatment coefficients and a Wald test of the joint significance of all treatments. Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

K. Survey Experiments

We conducted nationally representative online survey experiments in the United Kingdom, Poland, and the United States to investigate self-reported motives for deciding to return or keep a lost wallet. We conducted surveys in the UK and US in English. For Poland, we hired two professional translators—one for the Polish translation and the other to translate it back to English. We did this to ensure that the meaning of the questions were not lost in translation.

We sampled a total of 2,525 respondents through a Qualtrics online sample ($n = 829$ in the UK, $n = 809$ in Poland, and $n = 887$ in the US). To qualify for participation, individuals had to pass a simple attention check and meet the demographic quotas (based on age, gender, and residence) set by Qualtrics to construct the representative samples. Participants received a flat payment of US \$4.00 for their participation.

We randomly assigned participants to one of our four treatments corresponding to the NoMoney, Money, BigMoney, and Money-NoKey condition. Participants were told the study was about lost and found property, and then asked to rate their knowledge of lost property laws. They then read a brief description of a typical drop-off scenario and viewed a picture of the wallet and its contents. The particular description and picture of the wallet varied according to the condition. We also randomized the owner's name and the type of institution. Figure A2.3 provides an example of how this information was presented to participants.

After reading the scenario, participants completed several blocks of questions. In the first block, participants were asked how likely was it they would receive a financial reward from the owner if they were to contact him about the wallet, and responded on an 11-point scale from 0% to 100% in 10% increments. They were then asked, assuming the owner offered a financial reward, how much money they thought the owner would give. Participants provided their response in an open-text field.

In the second block, participants were asked the following questions on 11-point scales (0 = *not at all*, 10 = *very much*): “How concerned would you be with other people’s impression of you if you do not contact the owner?”, “How important do you think is the lost item for its owner?”, “To what extent would it feel like stealing if you do not contact the owner?”, and “How

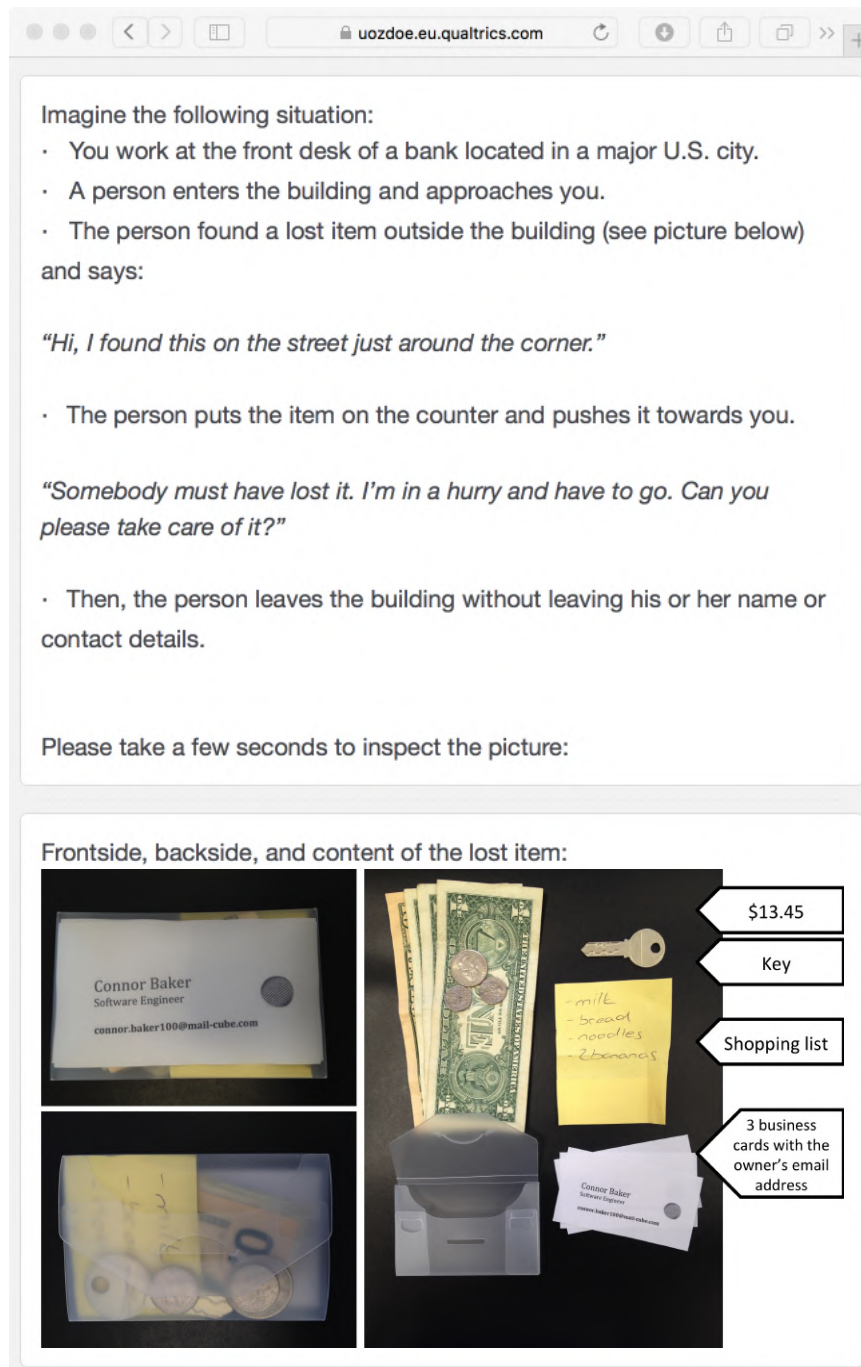


FIGURE A2.3
Example Survey Scenario

Notes. Scenarios and pictures were adjusted according to experimental condition and country. We also randomly varied the owner's name and type of institution in the scenario.

concerned would you be that you get punished if you do not contact the owner?" The order of questions in this block were randomized for each participant.

In the third block, participants were asked to guess the annual income of the owner compared to the average person in their country on a seven-point scale ($-3 = \textit{much lower than the average person}$, $+3 = \textit{much higher than the average person}$). In the fourth block, participants were asked how likely they would be to contact the owner to return the lost item, and also how likely that someone else would contact the owner to return the lost item in such a situation. For both questions participants responded on an 11-point scale from 0% to 100% in 10% increments.

We then included a number of exploratory questions. Participants were asked if they personally have ever lost a wallet, a mobile phone, or a key, as well as if they have ever found a lost wallet, mobile phone, or key. For each item they responded either yes (coded as 1) or no (coded as 0). In another block participants completed seven items from the Empathic Concern subscale of the Interpersonal Reactivity Index (Davis, 1983). Participants also completed six items from the Impression Management subscale of the Balanced Inventory of Desirable Responding (Paulhus, 1984), and a four-item measure of general attitudes about honesty. Lastly, participants were presented with several misbehaviors (e.g., cheating on one's taxes) and asked to assess the degree that most other people would consider the behavior appropriate or inappropriate on a four-point scale ($-2 = \textit{very inappropriate}$, $+2 = \textit{very appropriate}$).

As an additional attention check, we asked participants to recall key details from the study. We first asked them to list the contents of the wallet in a series of open-text boxes. We then asked them to identify the name of the owner from a list of 6 options. Finally, we asked them to recall the amount of money in the wallet in an open-text box.

L. Prediction Study: Non-Expert Sample

We conducted an online survey in the United States to investigate lay beliefs about the relationship between civic honesty and monetary incentives. Our sample consisted of 299 U.S. adults from Amazon.com's Mechanical Turk labor market (58% male, 42% female; M age = 35.49, SD = 10.66). To qualify for participation, individuals had to take the survey using a non-mobile device (such as a desktop or laptop computer) and pass a simple attention check. Participants received a flat payment of US \$0.50 for their participation, along with the opportunity to win a \$5.00 bonus.

Participants were told that we had recently conducted a study in 25 US cities, and their job was to predict the outcomes of the study. We first described the general design of our lost wallet experiments, then provided participants with details about the exact procedure, the wallets we turned in, and details about three of our experimental conditions (NoMoney, Money, and BigMoney). Participants were also provided with an image of the wallets similar to that in Figure A2.3. We then asked participants to predict reporting rates (from 0-100%) for each condition. We informed participants that they should try their best to be accurate, as the most accurate 5% of participants in the study would receive a bonus payment of \$5.00. All responses were made on the same page using slider scales from 0 to 100.

On the next page we probed participants' beliefs about the relevant motivations operating in each of our experimental conditions. We first asked participants to consider the following three issues that our recipients may have been considering when deciding to return or not return a wallet: (i) how tempted would the recipient be to keep the money in the wallet, (ii) how concerned would the recipient be for the owner, and (iii) how much would the recipient feel like they were a thief if they did not return the wallet. Participants estimated the relative importance of these three concerns for each condition on 100-point slider scales, with higher numbers indicated greater importance. For each condition, responses for the three concerns were required to sum to 100.

Afterwards, participants provided basic demographic information including their age, gender, ethnicity, educational level, employment status, and household income.

M. Prediction Study: Expert Sample

We conducted a follow-up online survey to investigate expert predictions about the relationship between civic honesty and monetary incentives. To do so, we surveyed a group of academic economists whose email addresses were publicly available on the Research Papers in Economics repository website (<http://repec.org>).

We culled email addresses for economists who have published in the last five years, and who ranked in the top 5% in at least one of the following dimensions on the website: “average rank,” “citations,” “citations, discounted by age,” “h-index,” “abstract views,” and “downloads.” To exclude economists who were likely to be familiar with our project, we excluded anyone from our email list who was affiliated with a research institute in Zurich or on the website’s expert lists for experimental economics, cognitive and behavioral economics, norms and social capital, or prospect theory. This procedure yielded 2,283 email addresses. We sent out an invitation to participate in the study, and received 294 completed responses. For our analysis we excluded 15 respondents who reported familiarity with our lost wallet experiments, yielding a final sample of 279 participants (88% male, 12% female; M age = 54.60, SD = 11.64). The overwhelming majority of respondents were university professors (95%), with 71% at the rank of full professor.

Participants were given the same instructions and were asked to make the same predictions as in our previous prediction study, but were not asked to complete the motivation items on self-interest, altruism, and theft aversion. Participants were informed up front that the three most accurate respondents would receive a US \$100 bonus which they could keep or donate to charity. At the end of the survey we asked respondents to report their gender, age, current academic status/ranking, and whether they were previously familiar with our lost wallet experiments.

II. A CONCEPTUAL FRAMEWORK FOR CIVIC HONESTY

We model a recipient's decision to return a lost wallet as follows. A recipient chooses an action $a \in \{0, 1\}$ to either keep the wallet ($a = 0$) or return the wallet along with its content ($a = 1$). The recipient's decision is determined by four factors. The first factor reflects the effort necessary to return the wallet. The recipient incurs an effort cost c when returning the wallet, such as the time required to contact the owner. The second factor reflects the potential material benefits to the recipient. If the recipient decides to keep the wallet, then her material payoff increases by the amount of money m in the wallet. The third factor reflects potential altruistic concerns from the recipient towards the owner, captured by the weight α that the recipient places on the potential externality. If the recipient fails to return the wallet then she can internalize the costs to the owner, which includes the money inside the wallet (m) along with anything else inside the wallet thought to be valuable to the owner (v). Based on prior empirical work (Engel, 2011; Andreoni and Miller, 2002; Charness and Rabin, 2002), we assume that the recipient cannot value the wallet more than its owner ($0 \leq \alpha < 1$). The fourth factor reflects self-image concerns, captured by the weight γ (hereafter what we call "theft aversion"). If the recipient fails to return the wallet then she may consume a negative self-image resulting from thinking of herself as a dishonest person. The weight placed on theft aversion is assumed to be non-negative, $\gamma \geq 0$. Based on these four factors, an individual chooses action a in order to maximize the following objective function:

$$\max_{a \in \{0, 1\}} \{(1 - a)m + a\alpha(m + v) - (1 - a)\gamma m - ac\}. \quad (\text{A2.2})$$

As is clear from equation (A2.2), we assume the non-pecuniary costs of failing to return the wallet (captured by α and γ) increase linearly with the amount of money inside the wallet.⁶ It

6. This is a reduced form representation consistent with signaling models such as Bénabou and Tirole (2006), where recipients are concerned about their social or self-image. Returning a wallet with greater amounts of money is a costlier signal about the recipient's honesty and therefore yields a higher reputational benefit than a wallet with smaller amounts of cash. Psychological costs could also be represented in other forms, such as negative emotional costs (Charness and Dufwenberg, 2006; Battigalli and Dufwenberg, 2007) or a desire to adhere to social norms (Krupka and Weber, 2013; Andreoni and Bernheim, 2009).

follows from equation (A2.2) that a recipient will return a wallet if and only if

$$\alpha v + m(\alpha + \gamma - 1) \geq c. \quad (\text{A2.3})$$

Note that in our framework theft aversion depends on the amount of money in the wallet, whereas altruistic concerns for the owner depend on the amount of money as well other contents in the wallet thought to be valuable to the owner. Accordingly, recipients sufficiently high in altruism (i.e., a high α) would be compelled to return the wallet even when it contains little or no money. By contrast, recipients who are theft averse (i.e., high γ) would be compelled to return a wallet only when it contains sufficiently large amounts of money.

When we apply the framework to our current experiments, we obtain four potential types of recipients. The first type involves recipients primarily motivated by material self-interest (i.e., low α and low γ), who will never return a wallet regardless of its contents. Our second type involves recipients who are sufficiently altruistic and theft averse (i.e., high α and high γ) who will always return the wallet regardless of its contents (so long as such concerns outweigh the effort costs of returning the wallet).

The third and fourth types are unique in that their behavior will depend on the wallet's contents. Our third type involves recipients high in altruism but low in theft aversion (i.e., high α and low γ), who will return a wallet with little to no money but will keep a wallet when it contains sufficiently large amounts of cash. These individuals are primarily motivated by altruistic concerns for low amounts of money, but self-interest dominates for larger amounts of money. Formally this type is comprised of individuals where

$$\frac{c + m(1 - \gamma)}{v + m} > \alpha > \frac{c}{v}. \quad (\text{A2.4})$$

Our fourth type involves recipients low in altruism but high in theft aversion (i.e., low α and high γ), who will fail to return a wallet with little to no money but return a wallet when it contains larger amounts of cash. These individuals will not be sufficiently motivated by concern for the owner's welfare to return wallets with relatively small amounts of money, but theft

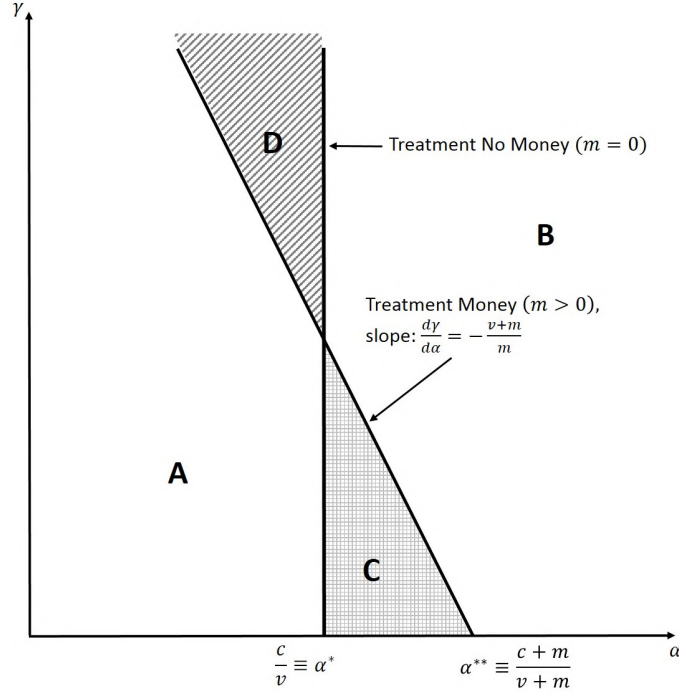


FIGURE A2.4

Response Patterns as a Function of Altruism (α) and Theft Aversion (γ)

Notes. This figure illustrates response patterns for each of four behavioral types as a function of altruism (α) and theft aversion (γ). Recipients in region A will not report a wallet in either treatment. In contrast, recipients in region B will always report a wallet, regardless of whether it contains money or not. Recipients in region C are sufficiently altruistic to return a wallet in the NoMoney condition, but their degree of theft aversion is not large enough to compensate the temptation to pocket the money in Money condition. Finally, region D comprises recipients who are not sufficiently altruistic to report a wallet in the Money condition, but their degree of theft aversion is strong enough to induce them to return the wallet in the NoMoney condition.

aversion concerns will dominate for larger amounts of money. Formally this type is comprised of individuals where

$$\frac{c}{v} > \alpha > \frac{c + m(1 - \gamma)}{v + m}. \quad (\text{A2.5})$$

The distribution of types in the population determines the nature of the relationship between the reporting rate and the amount of money in the wallet. Figure A2.4 illustrates this dynamic along a α/γ -plane for the NoMoney and Money conditions. In the NoMoney condition, recipients with sufficiently high altruistic concerns ($\alpha > \alpha^*$) will return the wallet, while all other recipients will not; the separation between these two response types is denoted by the vertical line in Figure A2.4. In the Money condition, recipient types are distinguished by the line with slope $-(v + m)/m$ which intersects the α -axis to the right of α^* at $\alpha^{**} = (c + m)/(v + m)$.

The two lines divide the plane into four regions. Recipients in region A fail to return the wallet in both conditions because they are self-regarding, reflecting our first type (low α and low γ). Recipients in region B will return the wallet in both treatments because they are sufficiently altruistic and theft averse, reflecting our second type (high α and high γ). Region C consists of recipients who are altruistic enough to return the wallet in the NoMoney condition but fail to return the wallet in the Money condition due to self-interest, reflecting our third type (i.e., high α and low γ). Finally, region D consists of recipients who do not reach the threshold of altruism α^* in the NoMoney condition and therefore do not return the wallet, but who are sufficiently motivated by theft aversion to return the wallet in the Money condition.

Based on our framework, treatment differences in reporting rates reflect the distribution of types in the population. The fact that reporting rates are relatively higher in the Money condition suggest that recipient types in region D are more prevalent than those in region C. An analogous line of reasoning can be applied to explain the increase in civic honesty in the BigMoney condition relative to the NoMoney and Money conditions.

III. RESULTS

A. Civic Honesty across Countries

We first examine reporting rates in the NoMoney and Money conditions for all 40 countries. Overall, 51% of recipients in the Money condition reported the wallet compared to 40% of recipients in the NoMoney condition ($Z = 14.18$, $P < 0.0001$). We observe an increase in reporting rates for the Money condition relative to the NoMoney condition in 38 out of 40 countries, and this effect is statistically significant at the 5% level for 19 countries after adjusting for the pairwise comparison false discovery rate (Benjamini and Hochberg, 1995). Furthermore, in neither of the two countries that displayed a reduction in reporting rates in the Money condition was the decline statistically significant ($Z = 1.47$, $P = 0.141$ for Mexico; $Z = 0.19$, $P = 0.853$ for Peru).

Table A2.8 displays the results when aggregated across all 40 countries. For the table as well as all subsequent analyses, we use ordinary least squares (OLS) regression with robust standard errors. Responses are coded as 100 if the wallet was reported and 0 otherwise. We use OLS for purposes of simplicity and clarity because coefficients can be directly interpreted as percentage point changes; using nonlinear models such as logistic regression return virtually identical results. Column (1) of Table A2.8 indicates that reporting rates increase by 10.8 percentage points in the Money relative to the NoMoney condition when including city, institution, and treatment fixed effects⁷ ($t_{16941} = 15.16$, $P < 0.001$).

Column (2) of Table A2.8 indicates that our treatment effect holds when also controlling for additional recipient and situational characteristics. This specification also finds that these additional characteristics also influenced reporting rates independent of our experimental conditions. On average men were roughly 2 percentage points less likely than women to report a wallet ($t_{16928} = 2.78$, $P = 0.005$), and older recipients (i.e., those judged to 40 years or older) were 2 percentage points less likely to report a wallet ($t_{16928} = 2.75$, $P = 0.006$). The presence of a computer at the recipient's workstation increased the likelihood of reporting a wallet ($t_{16928} = 7.10$,

7. Controlling for the other two experimental conditions does not affect estimates of the Money coefficient, but provides added precision when estimating our other control variables.

TABLE A2.8
REPORTING RATES IN THE MONEY AND NOMONEY CONDITION

	(1)	(2)
Money	10.828*** (0.714)	10.792*** (0.712)
Male		-2.076** (0.747)
Age 40+		-2.030** (0.738)
Computer		6.874*** (0.969)
Coworkers		4.675*** (0.765)
Other bystanders		-3.900*** (0.795)
Constant	34.620** (11.434)	33.302** (11.112)
Controls:		
Institution FE	yes	yes
City FE	yes	yes
Treatments	yes	yes
Observations	17303	17295
Adjusted R^2	0.178	0.185

Notes. OLS estimates with robust standard errors in parentheses. The dependent variable in all models takes on the value 100 if a wallet was reported and 0 otherwise. “Money” is a dummy for treatment Money (we also include an indicator for treatments “BigMoney” and “Money-NoKey”). The omitted category in this table is the treatment “NoMoney.” All models further include city and institution fixed effects. In Column (2), we also include binary control variables for individual and situational factors, including a recipient’s age (above 40 years) and gender (male), as well as the presence of a computer, coworkers, and other bystanders. Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

$P < 0.001$), as did the presence of other coworkers ($t_{16928} = 6.11$, $P < 0.001$). The latter of the two findings is unsurprising given that, in addition to the possibility of increased social monitoring, the presence of other coworkers may have also reduced recipients’ workload. By contrast, the presence of other bystanders (excluding coworkers) decreased reporting rates ($t_{16928} = 4.90$, $P < 0.001$). One possibility for this result is that the increase in workload by having bystanders present exerted a larger influence on recipients’ behavior than did the additional social pressure brought about by the bystander’s presence.

TABLE A2.9
REPORTING RATES IN NoMoney, Money, AND BigMoney CONDITION

	UK, Poland, and US	United Kingdom	Poland	United States
	(1)	(2)	(3)	(4)
Money	15.940*** (2.370)	23.106*** (3.851)	3.310 (4.690)	18.301*** (3.934)
BigMoney	25.235*** (2.558)	35.941*** (4.567)	11.761** (4.410)	27.832*** (4.260)
Constant	35.506*** (8.517)	25.763** (9.345)	59.380*** (11.216)	34.445** (11.291)
Controls:				
Recipient	yes	yes	yes	yes
Situation	yes	yes	yes	yes
Institution FE	yes	yes	yes	yes
City FE	yes	yes	yes	yes
Other treatments	yes	yes	yes	yes
Money = BigMoney	0.000	0.001	0.058	0.027
Observations	2926	1132	794	1000
Adjusted R^2	0.091	0.122	0.050	0.100

Notes. OLS estimates with robust standard errors in parentheses. Column (1) presents the results for all three countries, Column (2) for the United Kingdom, Column (3) for Poland, and Column (4) for the United States. The dependent variable in all models takes on the value 100 if a wallet is reported and 0 otherwise. “Money” and “BigMoney” are treatment indicators (we also include an indicator for our “Money-NoKey” treatment but report those estimates in Table A2.10). The omitted category in this table is the treatment “NoMoney.” All models include binary control variables for recipient and situational characteristics, including a recipient’s age (above 40 years) and gender (male), as well as the presence of a computer, other people, and coworkers. All models include city and institution fixed effects. The “Money = BigMoney” row reports P -values from t -tests for equality of the treatment coefficients. Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

B. Civic Honesty under High Stakes

We next examine reporting rates for the three countries in which we conducted the BigMoney condition alongside our Money and NoMoney conditions ($N = 2,932$). Despite the higher incentive to steal, recipients were more likely to report a lost wallet when it contained greater amounts of money. Across the three countries, 46% of the recipients reported the wallet in the NoMoney condition, which increased to 61% in the Money condition and increased even further to 72% in the BigMoney condition ($Z > 4.40$ for all pairwise comparisons, $P < 0.001$). Column (1) in Table A2.9 shows that, when controlling for situational and recipient characteristics, the average share of recipients who reports a wallet increases by almost 16 percentage points in the Money relative to the NoMoney condition ($t_{2846} = 6.73$, $P < 0.001$). The BigMoney condition increases the reporting rate by 25 percentage points, on average, relative to the NoMoney condition ($t_{2846} = 9.86$, $P < 0.001$), and the difference between the BigMoney and Money conditions is also significant ($t_{2846} = 3.92$, $P < 0.001$). Columns (2) to (4) show that the increasing trend

TABLE A2.10
REPORTING RATES IN MONEY-NOKEY CONDITION

	UK, Poland, and US	United Kingdom	Poland	United States
	(1)	(2)	(3)	(4)
Money-NoKey	-9.185*** (2.482)	-11.750** (3.832)	-9.820* (4.743)	-2.927 (4.433)
Constant	51.446*** (8.393)	48.869*** (8.971)	62.690*** (11.068)	52.746*** (11.373)
Controls:				
Recipient	yes	yes	yes	yes
Situation	yes	yes	yes	yes
Institution FE	yes	yes	yes	yes
City FE	yes	yes	yes	yes
Other treatments	yes	yes	yes	yes
Observations	2926	1132	794	1000
Adjusted R^2	0.091	0.122	0.050	0.100

Notes. OLS estimates with robust standard errors in parentheses. Column (1) presents the results for all three countries, Column (2) for the United Kingdom, Column (3) for Poland, and Column (4) for the United States. The dependent variable in all models takes on the value 100 if a wallet is reported and 0 otherwise. “Money-NoKey” is a treatment indicator (we also include indicators for treatments “NoMoney” and “BigMoney” but do not report their estimates for ease of exposition). The omitted category in this table is the treatment “Money.” All models include binary control variables for individual characteristics and situational factors, including a recipient’s age (above 40 years) and gender (male), as well as the presence of a computer, other people, and coworkers. All models include city and institution fixed effects. Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

in civic honesty for larger monetary stakes holds for all three countries.

C. Testing for Altruism

To examine the role of altruism, we compare the Money condition to the Money-NoKey condition for the three countries where we conducted both treatments ($N = 2,932$). Wallets from these two conditions contain the same contents with the exception of the key, which is valuable to the owner of the wallet but not to the recipient.⁸ As a result, altruistic concerns should be responsible for any differences in reporting rates between the Money and Money-NoKey conditions. Shown in Table A2.10, we do find relatively fewer wallets were reported when they did not contain a key. Column (1) indicates that the average reporting rate across countries decreased by more than 9 percentage points in the Money-NoKey condition relative

8. In the representative survey experiments, we asked participants to evaluate the importance of the wallet to the owner on a 11-point scale from not at all (0) to very much (10). Consequently, respondents tended to recognize the value of the key to the owner. On average, respondents considered the wallet in the Money-NoKey condition to be 2.32 points (or 0.86 standard deviations) less important to the owner compared than the wallet in the Money condition ($t_{1120} = 14.33$, $P < 0.001$). This comparison is in the same direction and statistically significant when examining each country separately (all P -values < 0.001).

to the Money condition ($t_{2846} = 3.70$, $P < 0.001$). Columns (2) to (4) show that this pattern holds for all three countries, though the difference was statistically significant only for the UK and Poland (12 and 10 percentage points, respectively).

D. Evidence for Theft Aversion

Since we collected survey data to measure how possible psychological motives to report a lost wallet differ according to wallet content, we restrict our analysis to participants who were able to correctly recall the amount of money inside the wallet described to them (rounded to the nearest integer). This leaves us with a sample of 2,160 participants from our original sample of 2,525. When we do not exclude any participants we find largely similar results (displayed in Table A2.12) to those reported below.

In our survey experiments, we asked participants to rate the extent to which failing to report a wallet felt like stealing. Column (1) in Table A2.11 shows that across the three countries, respondents reported that failing to return a wallet would feel more like stealing when the wallet contained greater amounts of money. Relative to the NoMoney condition, the average score increased by 1.57 points (or 0.47 standard deviations) in the Money condition, and by 2.08 points (or 0.64 standard deviations) in the BigMoney condition ($t_{2150} = 7.72$, $P < 0.001$ for Money; $t_{2150} = 10.41$, $P < 0.001$ for BigMoney). The difference between the Money and BigMoney condition was also significant ($t_{2150} = 2.71$, $P = 0.007$). In contrast, we failed to observe a reliable difference in responses between the Money and Money-NoKey conditions ($t_{2150} = 1.13$, $P = 0.259$). This suggests that anticipated costs due to theft aversion depend on the amount of money in the wallet, but do not meaningfully depend on other contents that are only valuable to the owner.

In the survey we also asked respondents to report the likelihood they would contact the owner to return the wallet (from 0-100%). Naturally such self-reports should be interpreted with caution, and indeed we find responses were overly optimistic when compared with the behavioral data (average estimates ranged between 88% and 93% across countries). Nevertheless, the pattern of treatment differences in self-reported likelihood of returning wallet follow the same rank-ordering as those from our lost wallet experiments (see Column 2 in Table A2.11), and so

TABLE A2.11
SURVEY RESPONSES ACROSS EXPERIMENTAL CONDITIONS

	Theft aversion concerns	Stated likelihood of reporting (in %)		
	(1)	(2)	(3)	(4)
Money	1.570*** (0.203)	2.400* (0.985)	-0.748 (0.966)	-0.368 (0.941)
BigMoney	2.076*** (0.200)	3.847*** (0.975)	-0.315 (0.989)	-0.928 (0.979)
Money-NoKey	1.358*** (0.201)	-2.454* (1.171)	-5.177*** (1.131)	-1.843 (1.163)
Theft aversion concerns			2.005*** (0.161)	1.690*** (0.152)
Perceived importance to owner				1.283*** (0.180)
Fear of punishment				0.133 (0.106)
Constant	6.512*** (0.224)	86.414*** (1.217)	73.357*** (1.675)	65.609*** (2.126)
Controls:				
Institution FE	yes	yes	yes	yes
Country FE	yes	yes	yes	yes
Money = BigMoney	0.007	0.120	0.623	0.516
Money = Money-NoKey	0.259	0.000	0.000	0.165
Observations	2160	2160	2160	2160
Adjusted R^2	0.053	0.029	0.159	0.188

Notes. OLS estimates with robust standard errors in parentheses. In Column (1) the dependent variable is our proxy for theft aversion concerns which is measured by the question “To what extent would it feel like stealing if you do not contact the owner?” with possible answers ranging from “not at all” (0) to “very much” (10). The dependent variable in Columns (2) to (4) is the likelihood that participants would report the wallet (as a percentage). “Money,” “BigMoney,” and “Money-NoKey” are treatment indicators. All models include country and institution fixed effects. The bottom of the table reports P -values from t -tests for equality of the treatment coefficients. Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

we use our self-report data as a proxy for exploring possible motives for returning a lost wallet.

Column (3) in Table A2.11 shows that theft aversion concerns were positively related to one’s stated likelihood of reporting a wallet ($t_{2149} = 12.44$, $P < 0.001$). Furthermore, compared to the model displayed in Column (2) that does not control for theft aversion, the model in Column (3) provides a substantially better fit to the data (adjusted R^2 increases from 0.029 to 0.159) and the coefficients for the Money and BigMoney conditions shrink and are no longer statistically significant. To the extent such self-reports extend to real behavior, theft aversion may partly explain why people are more likely to return a lost wallet with greater amounts of money inside. Finally, Column (4) also includes a measure of perceived importance of the wallet to the owner, which serves as a proxy for altruistic concerns, and a measure for the subjective fear

TABLE A2.12
SURVEY RESPONSES ACROSS EXPERIMENTAL CONDITIONS, FULL SAMPLE

	Theft aversion concerns	Stated likelihood of reporting (in %)		
	(1)	(2)	(3)	(4)
Money	1.501*** (0.189)	2.919** (0.958)	-0.121 (0.939)	0.219 (0.916)
BigMoney	1.742*** (0.190)	3.708*** (0.968)	0.181 (0.975)	-0.423 (0.960)
Money-NoKey	1.225*** (0.189)	-1.736 (1.124)	-4.217*** (1.082)	-0.786 (1.112)
Theft aversion concerns			2.025*** (0.151)	1.699*** (0.140)
Perceived importance to owner				1.436*** (0.172)
Fear of punishment				0.059 (0.099)
Constant	6.653*** (0.208)	85.616*** (1.158)	72.143*** (1.573)	63.917*** (2.003)
Controls:				
Institution FE	yes	yes	yes	yes
Country FE	yes	yes	yes	yes
Money = BigMoney	0.167	0.371	0.716	0.427
Money = Money-NoKey	0.110	0.000	0.000	0.312
Observations	2525	2525	2525	2525
Adjusted R^2	0.039	0.023	0.152	0.185

Notes. OLS estimates with robust standard errors in parentheses. Full sample, including participants that failed our recall attention check. In Column (1) the dependent variable is our proxy for theft aversion concerns measured by the question “To what extent would it feel like stealing if you do not contact the owner?” with possible answers ranging from “not at all” (0) to “very much” (10). The dependent variable in Columns (2) to (4) is the likelihood that participants would report the wallet (as a percentage). “Money,” “BigMoney,” and “Money-NoKey” are treatment indicators. All models include country and institution fixed effects. The bottom of the table reports P -values from t -tests for equality of the treatment coefficients. Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

of punishment if the wallet is not reported. We find that both the perceived importance of the wallet and the aversion to viewing oneself as a thief are positively related to the stated likelihood of reporting the wallet ($t_{2147} = 7.14$, $P < 0.001$ and $t_{2147} = 11.11$, $P < 0.001$, respectively). This suggests that both altruism and theft aversion concerns are relevant to reporting a lost wallet, and that the two operate independently of each other. In contrast, self-reported fear of punishment is not significantly correlated with the stated likelihood of reporting the wallet ($t_{2147} = 1.26$, $P = 0.208$). Thus, if anything, threat of punishment plays only a minor role in reporting a lost wallet.

The pattern of results displayed in Table A2.11 suggests that theft aversion explains why

the reporting rate increases with the amount of money in the wallet, but not with the presence or absence of the key. To test this hypothesis, we conducted a series of mediation analyses. For the first mediation test, we restricted observations to the three conditions that only varied the amount of money in the wallet (NoMoney, Money, and BigMoney conditions). Using the NoMoney condition as our reference variable, we calculated indirect paths {Money \rightarrow theft aversion \rightarrow Likelihood of reporting} and {BigMoney \rightarrow theft aversion \rightarrow Likelihood of reporting} using bootstrapped standard errors with 10,000 resamples (Shrout and Bolger, 2002). Consistent with our hypothesis, theft aversion mediated the relationship between the amount of money inside the wallet and the likelihood of reporting a lost wallet (indirect $b_{Money} = 2.47$, $SE = 0.42$, $P < 0.001$; indirect $b_{BigMoney} = 3.26$, $SE = 0.48$, $P < 0.001$). Furthermore, the direct effect in both conditions was nonsignificant after accounting for the indirect effect of theft aversion (direct $b_{Money} = 0.003$, $SE = 0.97$, $P = 0.997$; direct $b_{BigMoney} = 0.43$, $SE = 0.99$, $P = 0.666$).

We then conducted a second mediation test based on our framework's assumption that altruism, rather than theft aversion, should explain the difference in reporting rates between the Money and Money-NoKey conditions. Restricting observations to only those two conditions, we conducted a similar path analysis as before except this time for the indirect paths {Money-NoKey \rightarrow Perceived harm to owner \rightarrow Likelihood of reporting} and {Money-NoKey \rightarrow Theft aversion \rightarrow Likelihood of reporting}. Consistent with our conceptual framework, we find that our proxy for altruistic concerns (perceived harm to the owner) reliably mediates the difference between the two conditions (indirect $b = -3.58$, $SE = 0.57$, $P < 0.001$) while theft aversion does not (indirect $b = -0.45$, $SE = 0.41$, $P = 0.274$). Furthermore, the direct effect of experimental condition was nonsignificant after accounting for our indirect effects (direct $b = -0.96$, $SE = 1.10$, $P = 0.381$). Taken together these results are consistent with the hypothesis that theft aversion explain why the reporting rate increases with the amount of money in the wallet, but does not explain why the reporting rate decreases with the absence the key.

E. Prediction Data: Non-Expert Sample

We examined whether people anticipate our behavioral results by asking online participants to predict reporting rates in the US for wallets that contained \$0, \$13.45, and \$94.15. Contrary to the behavioral data, respondents predicted that reporting would be highest when the wallet contained no money ($M = 72.71$, $SD = 29.47$), lower when the wallet contained a modest amount of money ($M = 65.04$, $SD = 24.01$), and lower still when the wallet contained a substantial amount of money ($M = 54.55$, $SD = 28.88$). All three predictions were reliably different from one another (Table A2.13, Column 1; $t_{298} \geq 6.40$, $P < 0.001$ for all pairwise comparisons). For each condition we also compared the average predicted change to the actual change in reporting rates. The predicted change in reporting rates was always lower (i.e., more cynical) than the actual change in reporting rates ($t_{298} \geq 12.16$, $P < 0.001$ for all pairwise comparisons).

We next examined response profiles within participants.⁹ As the amount of money inside the wallet increased, 64% predicted a monotonic decrease in civic honesty, 18% predicted a monotonic increase in civic honesty, 3% predicted no change, and 15% displayed non-monotonic predictions. Using a sign test (coded as -1 = predicted a decrease in civic honesty, $+1$ = predicted an increase in civic honesty, 0 = all remaining responses), we find that reliably more participants expected rates of civic honesty to decrease than increase as wallet amounts became larger ($P < 0.001$).

Participants also reported their beliefs about the relative share of different motivations operating in each condition. On average participants expected self-interest to grow and altruistic concerns to shrink for wallets containing relatively more money. Compared to the NoMoney condition, participants expected the temptation of recipients to pocket the money to increase by 18.95 points (or 0.93 standard deviations) in the Money condition, and by 36.98 points (or 1.26 standard deviations) in the BigMoney condition ($t_{298} > 16.00$, $P < 0.001$ for both comparisons). The difference between the Money and BigMoney conditions was also significant ($t_{298} = 14.15$, $P < 0.001$). We see the reverse pattern for beliefs about altruistic concerns by recipients towards

9. We assume weak monotonicity when calculating percentages for response profiles. Results from our sign-tests do not meaningfully change when response profiles are instead calculated assuming strong monotonicity.

the owner of the wallet. Relative to the NoMoney condition, participants expected altruistic concerns to decrease by 23.95 points (or 0.92 standard deviations) in the Money condition, and by 42.15 points (or 1.31 standard deviations) in the BigMoney condition ($t_{298} > 15.80$, $P < 0.001$ for both comparisons). The difference between the Money and BigMoney conditions was also significant ($t_{298} = 14.96$, $P < 0.001$).

Recall that in the behavioral and self-report data, theft aversion appeared to play an important role in explaining variation across conditions appears. Respondents in our prediction study, on the other hand, afforded considerably less importance to concerns of theft aversion. Relative to the NoMoney condition, participants did expect concerns about viewing oneself as a thief to increase by 5 points (or 0.26 standard deviations) in the Money condition, and by 5.17 points (or 0.21 standard deviations) in the BigMoney condition ($t_{298} > 3.60$, $P < 0.001$ for both comparisons). The difference between the Money and BigMoney conditions was not statistically reliable ($t_{298} = 0.16$, $P = 0.875$). We also note differences in predicted theft aversion concerns across conditions were considerably smaller than those observed for predicted self-interest or altruism.

Lastly, we examined how inferences about motivations related to predictions of rates of civic honesty (Columns 2 to 4, Table A2.13). Self-interest scores were inversely related to predicted reporting rates (Column 2; $t_{298} = 9.54$, $P < 0.001$), and altruism scores were positively related to predicted reporting rates (Column 3; $t_{298} = 6.53$, $P < 0.001$). In both cases, the adjusted R^2 increases by more than a factor of two relative to our baseline model in Column (1), and the coefficients for our treatment coefficients shrink and are no longer statistically significant. However, as displayed in Column (4), theft aversion concerns were not reliably associated with predicted reporting rates ($t_{298} = 1.36$, $P = 0.174$). When compared to our baseline model, including theft aversion concerns in the model does not meaningfully increase explained variance and our treatment coefficients do not decrease in size.

The pattern of results displayed in Table A2.13 suggest that respondents' inferences about self-interest and altruism, but not concerns of theft aversion, underly their beliefs that response rates will decline for wallets with relatively more money. To test this hypothesis, we conducted

TABLE A2.13
PREDICTIONS OF REPORTING RATES ACROSS EXPERIMENTAL CONDITIONS

	(1)	(2)	(3)	(4)
Money	-7.672*** (1.199)	1.613 (1.634)	0.198 (1.729)	-8.120*** (1.272)
BigMoney	-18.164*** (2.297)	-0.041 (2.893)	-4.313 (2.965)	-18.627*** (2.300)
Self-interest		-0.490*** (0.051)		
Altruism			0.329*** (0.050)	
Theft aversion concerns				0.090 (0.066)
Constant	72.709*** (1.706)	77.533*** (1.716)	49.534*** (3.820)	70.950*** (2.180)
Money = BigMoney	0.000	0.327	0.007	0.000
Observations	299	299	299	299
Adjusted R^2	0.066	0.191	0.138	0.068

Notes. OLS estimates with participant-clustered standard errors in parentheses. The dependent variable is predicted reporting rates by recipients (from 0-100%). “Money” and “BigMoney” are treatment indicators. The omitted category is “NoMoney.” The bottom of the table reports P -values from t -tests for equality of the treatment coefficients. Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

a series of mediation analyses. For each our three motivation items,¹⁰ we calculated the indirect pathway {Experimental conditions \rightarrow Inferred motivation \rightarrow Predicted reporting rate} using bootstrapped participant-clustered standard errors with 10,000 resamples (Shrout and Bolger, 2002). Consistent with the pattern suggested in Table A2.13, inferences of increasing self-interest and declining altruism each statistically mediate the relationship between experimental conditions and predicted reporting rates (self-interest results: indirect $b_{Money} = -9.29$, SE = 1.18, $P < 0.001$; indirect $b_{BigMoney} = -18.12$, SE = 2.18, $P < 0.001$; altruism results: indirect $b_{Money} = -7.87$, SE = 1.29, $P < 0.001$; indirect $b_{BigMoney} = -13.85$, SE = 2.20, $P < 0.001$). However, we fail to observe a reliable indirect effect of inferred theft aversion concerns on predicted reporting rates (indirect $b_{Money} = 0.45$, SE = 0.36, $P = 0.209$; indirect $b_{BigMoney} = 0.46$, SE = 0.37, $P = 0.216$). Thus, participants appeared to weight the role of self-interest and declining altruism, but not inferences of theft aversion, in predicting rates of civic honesty.

10. We conducted separate mediation analyses for each motivation item rather than conduct a simultaneous mediation test for all three items, as the latter analysis would require us to remove at least one item due to collinearity (since inferences for the three items were required to sum to 100).

F. Prediction Data: Expert Sample

The results we observe from our expert sample were qualitatively similar to those from our MTurk sample, but considerably weaker in magnitude. On average, respondents predicted that reporting rates would be highest in the NoMoney condition ($M = 69.38$, $SD = 25.43$), followed by the Money condition ($M = 68.98$, $SD = 21.36$), and lowest in the BigMoney condition ($M = 65.70$, $SD = 23.15$). We compared conditions using an OLS regression with participant-clustered standard errors. Predicted reporting rates in the BigMoney condition were reliably lower than those in the NoMoney condition ($t_{278} = 2.05$, $P = 0.042$) and Money condition ($t_{278} = 2.50$, $P = 0.013$), but predicted reporting rates in the NoMoney and Money conditions did not reliably differ from one another ($t_{278} = 0.44$, $P = 0.660$). For each condition we also compared the average predicted change to the actual change in reporting rates. The predicted change in reporting rates was always lower (i.e., more cynical) than the actual change in reporting rates ($t_{278} \geq 8.70$, $P < 0.001$ for all pairwise comparisons).

We next examined response profiles within participants. As the amount of money inside the wallet increased, 49% predicted a monotonic decrease in civic honesty, 29% predicted a monotonic increase in civic honesty, 6% predicted no change, and 16% displayed non-monotonic predictions. Using a sign test (coded as -1 = predicted a decrease in civic honesty, $+1$ = predicted an increase in civic honesty, 0 = all remaining responses), we find that reliably more participants expected rates of civic honesty to decrease than increase as wallet amounts became larger ($P < 0.001$). In summary, experts in our sample held inaccurate beliefs, but to a lesser degree than our sample of MTurkers.

G. Cross-Country Correlates of Civic Honesty

In this section, we explore possible explanations for cross-country differences in civic honesty. To address potential issues related to reverse causality, we primarily consider “deep” and historical explanatory variables which are plausibly exogenous to honest behavior and are considered formative to the development of society (Spolaore and Wacziarg, 2013). To illustrate the value of this approach, consider that reporting rates in our study are positively correlated with contemporaneous measures of wealth (such as per capita income). From this correlation it is unclear whether country wealth leads to greater civic honesty or vice versa (or alternatively, some unobserved third variable influences both wealth and civic honesty). Now consider that, instead of wealth, we observed a correlation between a country’s geographic terrain and civic honesty. Country terrain can be considered a deep variable because civic honesty is unlikely to influence geography, but geography could potentially influence civic honesty (by shaping citizen’s interactions in ways that benefit or hinder cooperation). For this reason, using deep and historical variables is potentially more informative in explaining cross-country differences in civic honesty. We then extend our analysis to explore the role of culture and institutions, with the caveat that those factors may be endogenous¹¹ (Greif, 2006; Bisin and Verdier, 2011; Alesina and Giuliano, 2015; Enke, forthcoming).

We conducted a series of OLS regressions in which we regressed a given country-level variable onto individual decisions to report a wallet (for a full list of variables, see the “Country-level Correlates of Civic Honesty” subsection of Materials and Methods). As the rank-ordering of countries is almost identical for the NoMoney and the Money condition (Spearman’s $\rho = 0.939$, $P < 0.001$), we pooled data between the two conditions. All regressions control for treatment condition, recipient and situational characteristics, as well as institution fixed effects. Figure A2.5 presents the corresponding coefficients and standard errors (adjusted for clustering at the country-level). We standardized the explanatory variables to have a mean of zero and a

11. Some of the variables were not available for all countries in our dataset. Where possible, we updated the data to obtain better geographic coverage. For example, measures of historic institutions were substituted from predecessor countries and we manually coded linguistic traits for several countries using the *World Atlas of Language Structures (WALS)*. Figure A2.7 shows that the results are qualitatively similar if we only use data from the original sources.

standard deviation of one, so the coefficients can be interpreted as the difference in reporting rates associated with a one standard deviation change in the explanatory variable. To account for multiple hypothesis testing, we report P -values adjusted for the false discovery rate (Benjamini and Hochberg, 1995). Figs. A2.8 and A2.9 show that our results are robust when we conduct our regression analysis separately for the Money and NoMoney conditions.

We first examined whether rates of civic honesty are correlated with commonly-discussed geographic conditions: soil fertility, absolute latitude, distance to waterways, temperature, precipitation, elevation, and terrain ruggedness. These geographic conditions have been found to foster economic development (Ashraf and Galor, 2011b; Spolaore and Wacziarg, 2013), and we find that such variables are also significantly associated with civic honesty. Country-level reporting rates for lost wallets were associated with absolute latitude ($t_{39} = 5.26$, $P < 0.001$), lower temperature ($t_{39} = 4.40$, $P < 0.001$), and lower elevation ($t_{39} = 2.77$, $P = 0.020$). These findings suggest that civic honesty may be a channel through which geography affects economic development, in that geographic conditions and climate could have influenced the scope of social interactions and cooperation in pre-industrial societies. Norms of trust and cooperation may have in turn facilitated the transition from agricultural societies to market economies, which are based on interactions with out-group members and strangers (Ostrom, 1990; Greif, 1994; Woolcock, 1998; Henrich et al., 2001, 2010; Mokyr, 2008; Litina, 2016). Another possibility is that geography indirectly influences civic honesty by promoting favorable economic conditions, which in turn increases rates of honesty (Sharma et al., 2014; Fisman et al., 2015; Ananyev and Guriev, forthcoming).

We next examined the role of historical weather variability. Buggle and Durante (2017) advanced the hypothesis that subsistence farmers developed persistent norms of cooperation and trust in strangers to cope with climate risk, which in turn facilitated exchanges between communities or helped to establish geographically-diversified insurance agreements (Dean et al., 1985; Cashdan, 1985; Nettle, 1998). Using regional survey data from Europe, they found that historical weather variability is positively correlated with trust. Corroborating Buggle and Durante’s survey results, we find that historical seasonal variability in temperature is also positively

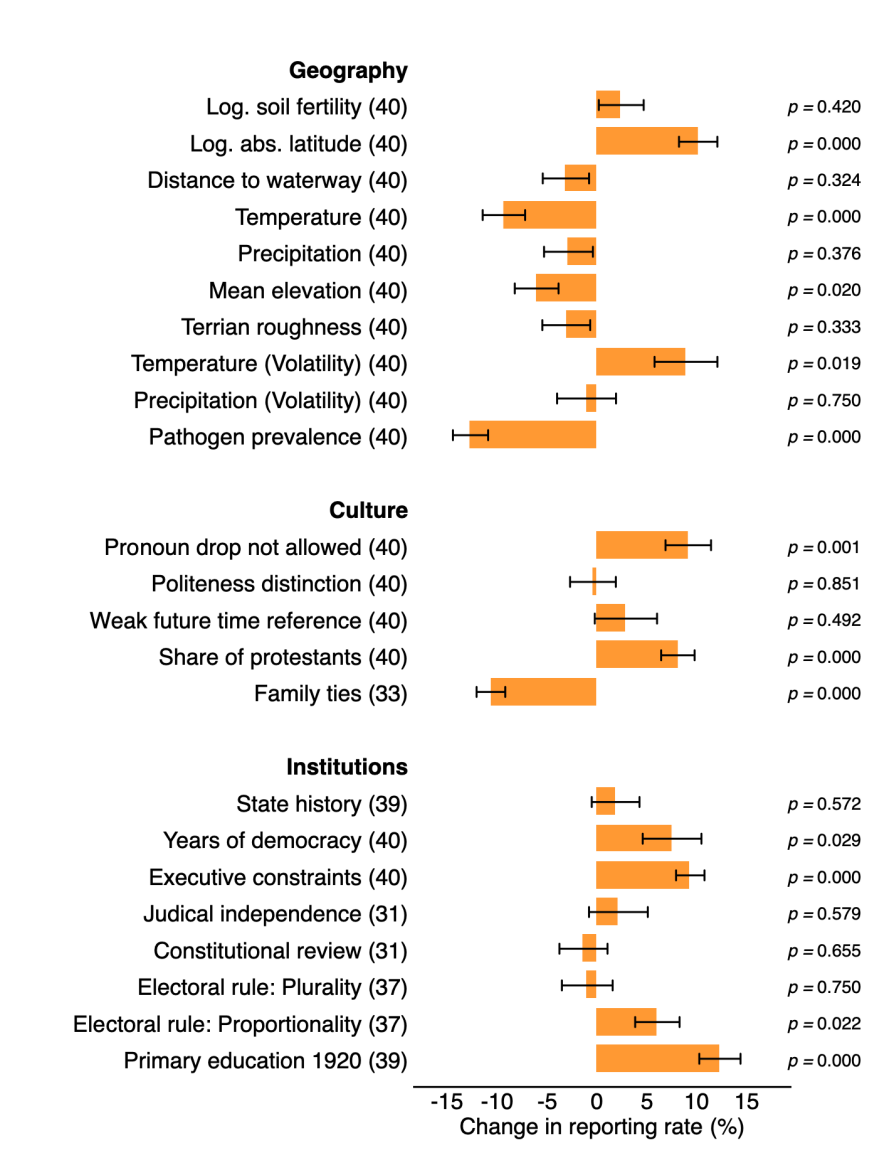


FIGURE A2.5
Correlates of Civic Honesty

Notes. OLS coefficient estimates with standard errors clustered at the country-level. The dependent variable takes on the value 100 when an individual reported a wallet and 0 otherwise. Each coefficient has been estimated separately using standardized explanatory variables. They can therefore be interpreted as the difference in reporting rates associated with a one standard deviation change in the explanatory variable. We control for treatment status, institution fixed effects, and our standard set of control variables for recipient and situational characteristics: dummies for age above 40 years and gender, as well as the presence of a computer, coworkers, and other bystanders. To correct for multiple hypothesis testing, *P*-values are adjusted for false discovery rate (Benjamini and Hochberg, 1995). The number of countries included in the regressions is indicated in parentheses.

correlated with reporting rates in our study ($t_{39} = 2.82$, $P = 0.019$).¹²

We conclude our analysis of geographic factors by examining the relationship between historical prevalence of infectious diseases and civic honesty. According to the prominent pathogen-stress theory of sociality, communities that lived in regions with high exposure to infectious diseases were less likely to interact with strangers to prevent potential infection of novel pathogens, and as a result adopted collectivistic norms limited to one’s immediate in-group (Fincher et al., 2008; Fincher and Thornhill, 2012). Given that the lost wallets in our study always belonged to a stranger, recipients in locations with historically high pathogen prevalence may have felt less compunction to return a lost wallet to an out-group member. Consistent with this hypothesis, we find a sizable negative association between historical pathogen prevalence and civic honesty ($t_{39} = 7.20$, $P < 0.001$).

We next explored the relationship between civic honesty and cultural proxies for a generalized sense of morality—that is, moral norms and obligations that extend beyond one’s in-group to anonymous strangers (Tabellini, 2010). To do so we first examined the role of different language structures, as language is thought to directly shape norms and expectations about behavior. For instance, Kashima and Kashima (1998) proposed that languages which do not permit the dropping of first person pronouns (e.g., “I” in English or “ich” in German) serve to demarcate an individual from his or her social context, in turn reinforcing values around individual autonomy and responsibility. We found a strong positive correlation between reporting rates and countries with languages which do not permit the dropping first personal pronouns ($t_{39} = 4.00$, $P < 0.001$). This finding is consistent with prior work demonstrating that individualistic values are positively related to behaviors in line with generalized morality norms (Tabellini, 2008a; Licht et al., 2007). By contrast, we failed to find a reliable correlation between reporting rates and the use of multiple second person pronouns ($t_{39} = 0.19$, $P = 0.851$) or weak future time reference ($t_{39} = 0.92$, $P = 0.492$), two linguistic features that have received attention in the literature.¹³

12. We also do not observe a significant correlation between precipitation and reporting rates ($t_{39} = 0.36$, $P = 0.750$). According to Galor and Savitskiy (2018), temperature shocks were more decisive for productivity in the pre-industrial era than precipitation.

13. Usage of multiple second person pronouns (e.g., “tu” and “vous” in French) as politeness distinction has been postulated to make status hierarchy and social distance more salient between speakers (Brown and Gilman,

Moving away from language to other cultural proxies of generalized morality, we next explored Protestantism. A long-standing literature in sociology and political science (Weber, 1930; Putnam et al., 1993) argues that Protestantism is conducive to social capital, and we find that countries with a higher share of Protestants also exhibit significantly more honest behavior ($t_{39} = 4.82$, $P < 0.001$). This is in line with prior work finding that Protestantism encourages applying the same behavioral standards to in-group and out-group members, leading to higher trust in strangers (Arruñada, 2010; La Porta et al., 1997; Glaeser et al., 2000; Uslaner, 2002; Guiso et al., 2011). Indeed, we also found that stronger family ties are negatively correlated with reporting rates ($t_{32} = 7.42$, $P < 0.001$), as stronger family ties imply norms of cooperation that are often limited to one's narrow in-group (Alesina and Giuliano, 2010; Banfield, 1958; Coleman, 1990; Alesina and Giuliano, 2014).

For the final part of our analysis, we explored some of the institutional determinants of civic honesty. The theoretical and empirical literature has examined both the complementarity between state formation and civic behavior (through the internalization of formal rules and increased trust in institutions), and their substitutability (as formal institutions may also crowd-out civic behavior) (Aghion et al., 2010; Bénabou and Tirole, 2011; Tabellini, 2008b; Cassar et al., 2014; Guiso et al., 2016; Lowes et al., 2017). We failed to find a significant association between state history—a commonly-used index of experience with formal government institutions (Bockstette et al., 2002)—and civic honesty ($t_{38} = 0.77$, $P = 0.572$). However, we found that both historical experience with democratic institutions and political constraints on executive power are positively correlated with reporting rates ($t_{39} = 2.55$, $P = 0.029$ for democratic history; $t_{39} = 6.54$, $P < 0.001$ for political constraints). This is consistent with the hypothesis that inclusive political institutions and the prevention of abuses of power are essential for civic behavior (Banfield, 1958; Putnam et al., 1993).

Some researchers have argued, however, that commonly-used measures of societal institutions are potentially problematic because they measure time-varying political outcomes rather than permanent constraints (Glaeser et al., 2004). To address this concern we also analyzed

1960; Kashima and Kashima, 1998). The weak future time reference feature allows the speaker to use the same grammatical tense to talk about future and present events and has been linked to greater patience and less impulsive behaviors (Chen, 2013; Falk et al., 2018).

a country's electoral rules (i.e., plurality and proportionality) and judicial checks and balances (i.e., judicial independence and constitutional review), which tend to be relatively time-invariant. Electoral systems based on plurality rule are thought to promote accountability due to the winner-take-all character of electoral competition,¹⁴ but at the cost of targeting benefits to narrow constituencies and less overall representativeness (Persson and Tabellini, 2004). Proportional representation, on the other hand, is thought to be more inclusive and promotes broader democratic consensus.¹⁵ Using data from Beck et al. (2001), we found that countries with proportional representation exhibit significantly higher reporting rates ($t_{36} = 2.71$, $P = 0.022$), while plurality representation is not reliably related to civic honesty ($t_{36} = 0.40$, $P = 0.750$). These results suggest that broad political representation could be a key factor underlying the correlation between democratic institutions and civic honesty. We also used judicial independence and constitutional review as constitutional measures of the judiciary's power to constrain the executive. While these measures have been associated with political and economic freedom in previous studies (La Porta et al., 1997), we failed to observe a significant correlation with reporting rates¹⁶ ($t_{30} = 0.72$, $P = 0.579$ for judicial independence; $t_{30} = 0.58$, 0.655 for constitutional review).

Our last institutional variable involves national education. The history of national education is closely intertwined with the formation of the modern state (Green, 1990; Uslander and Rothstein, 2016), so we examined the relationship between historical primary school enrollment rates and civic honesty. It has been argued that socialization is crucial to most primary education curricula and serves to ease interactions with strangers (Glaeser et al., 2007). We observed a significant and sizable positive correlation between historical rates of primary education and civic honesty ($t_{38} = 5.95$, $P < 0.001$), consistent with the hypothesis that education contributes to the formation of social capital (Knack and Keefer, 1997; Milligan et al., 2004; Helliwell and

14. The US and the UK are prime examples of countries with a plurality system where geographically defined constituencies elect one representative each.

15. Examples of proportional representation include Scandinavian countries where each constituency elects several representatives. In these countries additional mechanisms are in place to ensure that the allocation of seats closely mirrors the overall popular vote. However, plurality and proportional representation are not mutually exclusive. Elements of both systems can coexist if a country's constitution stipulates different rules for electing representatives in a two-chamber legislature (e.g., Switzerland) or if proportional representation is combined with some sort of bonus for the winning party, as is the case in Italy (Nannicini et al., 2013).

16. We note that for this analysis the sample is reduced to 31 countries due to data availability.

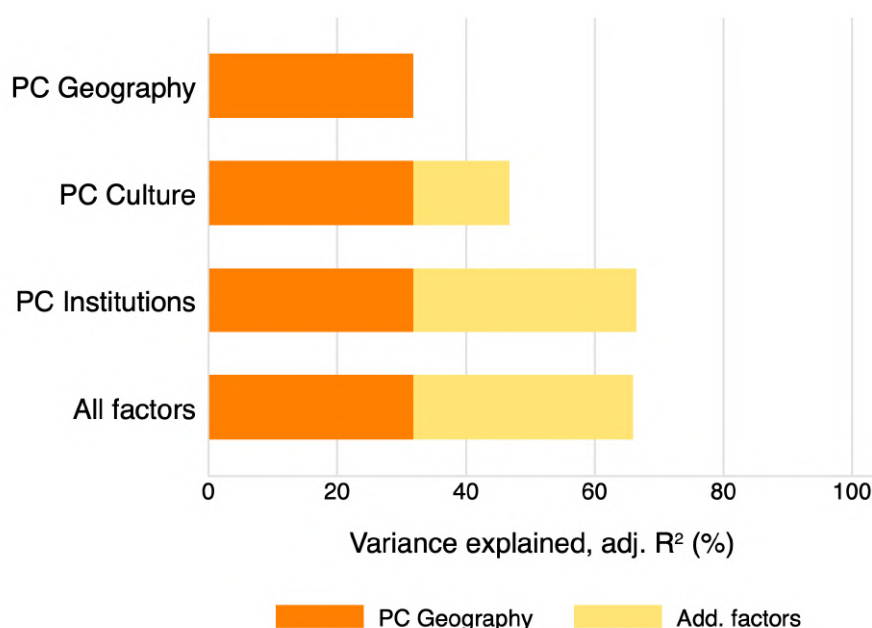


FIGURE A2.6
Explaining Cross-Country Variation

Notes. Explanatory power (adjusted R^2) of the first principal components of the geographic, cultural, and institutional variables. We regress country averages of regression-adjusted reporting rates (corrected for treatment indicators, institution fixed effects, and our standard set of control variables for individual characteristics and situational factors) on the first principal components of geography, geography and culture, geography and institutions, and all three categories together, respectively. To compute the first principal components of the variables in each category, we exclude variables with less than 37 observations (i.e., family ties, judicial independence, and constitutional review).

Putnam, 2007; Tabellini, 2010; Guiso et al., 2011; Algan and Cahuc, 2013).

Given that geography has been linked to culture and institutions (Engerman and Sokoloff, 1997; Acemoglu et al., 2002; Nunn and Wantchekon, 2011; Ashraf and Galor, 2011a; Alesina et al., 2013), it is possible that the correlations we observe between civic honesty and institutional variables may be spurious when not controlling for geographic conditions. We examined the robustness of our results to this concern by controlling for the first principal component of all geographic variables, and found qualitatively similar results¹⁷ (see Figure A2.10). The first principal component of our set of geographic variables accounts for roughly 32% of the variance in

17. Our results are similar if we control for the first three principal components or if the principal components are constructed using only the basic geographic factors, including soil fertility, absolute latitude, distance to waterway, temperature, precipitation elevation, and terrain ruggedness. As an alternative to controlling for the first principal component of geography, we also conducted the same regressions using the contemporary per capita income as our control variable. As shown in Figure A2.11, the results are largely unchanged.

civic honesty, and the first principal component of our set of cultural and institutional variables explains an additional 34% of the variation¹⁸ (Figure A2.6). Taken together, our analysis suggests that economically favorable geographic conditions, inclusive political institutions, national education, and cultural values that emphasize moral norms extending beyond one's in-group are positively associated with higher levels of civic honesty.

18. To compute the first principal components for each category, we exclude variables with less than 37 observations (i.e., family ties, judicial independence, and constitutional review). The results are similar if we restrict the analysis to the 25 countries where all measures are available: Geography explains 40% of the variation in civic honesty and culture and institutions together explain an additional 25%.

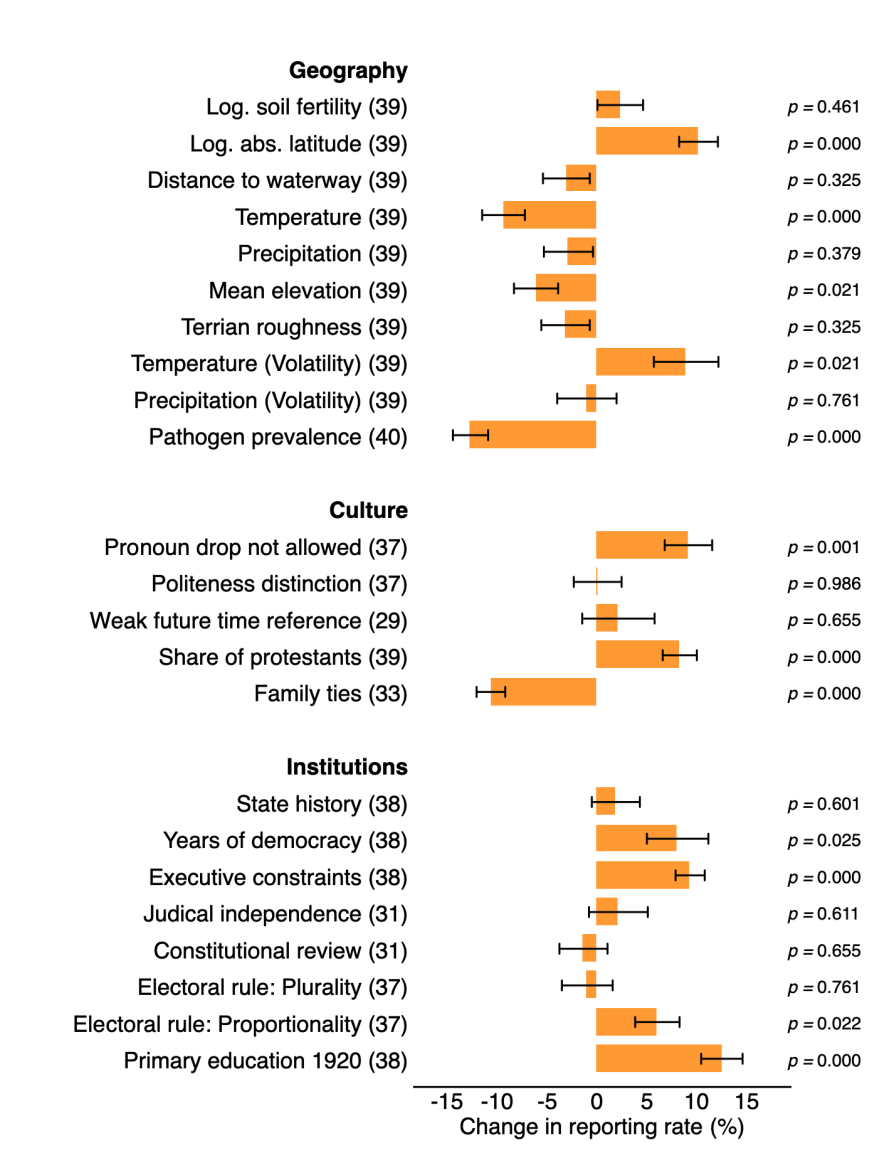


FIGURE A2.7

Correlates of Civic Honesty: Original Data Only

Notes. OLS coefficient estimates with standard errors clustered at the country-level. The dependent variable takes on the value 100 when an individual reported a wallet and 0 otherwise. Each coefficient has been estimated separately using standardized explanatory variables. They can therefore be interpreted as the difference in reporting rates associated with a one standard deviation change in the explanatory variable. We control for treatment status, institution fixed effects, and our standard set of control variables for recipient and situational characteristics: dummies for age above 40 years and gender, as well as the presence of a computer, coworkers, and other bystanders. To correct for multiple hypothesis testing, *P*-values are adjusted for false discovery rate (Benjamini and Hochberg, 1995). The number of countries included in the regressions is indicated in parentheses.

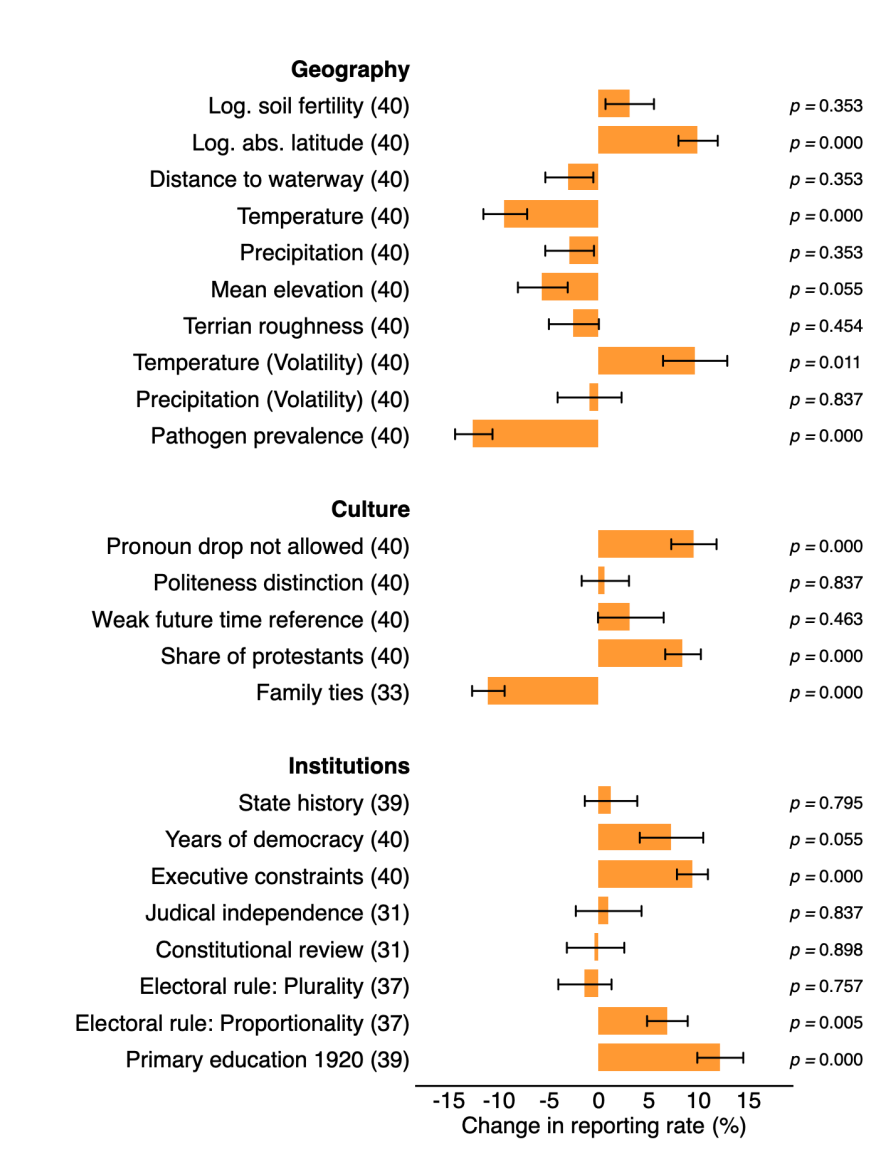


FIGURE A2.8
Correlates of Civic Honesty: NoMoney

Notes. OLS coefficient estimates with standard errors clustered at the country-level. The sample is restricted to drop-offs in treatment NoMoney. The dependent variable takes on the value 100 when an individual reported a wallet and 0 otherwise. Each coefficient has been estimated separately using standardized explanatory variables. They can therefore be interpreted as the difference in reporting rates associated with a one standard deviation change in the explanatory variable. We control for treatment status, institution fixed effects, and our standard set of control variables for recipient and situational characteristics: dummies for age above 40 years and gender, as well as the presence of a computer, coworkers, and other bystanders. To correct for multiple hypothesis testing, P -values are adjusted for false discovery rate (Benjamini and Hochberg, 1995). The number of countries included in the regressions is indicated in parentheses.

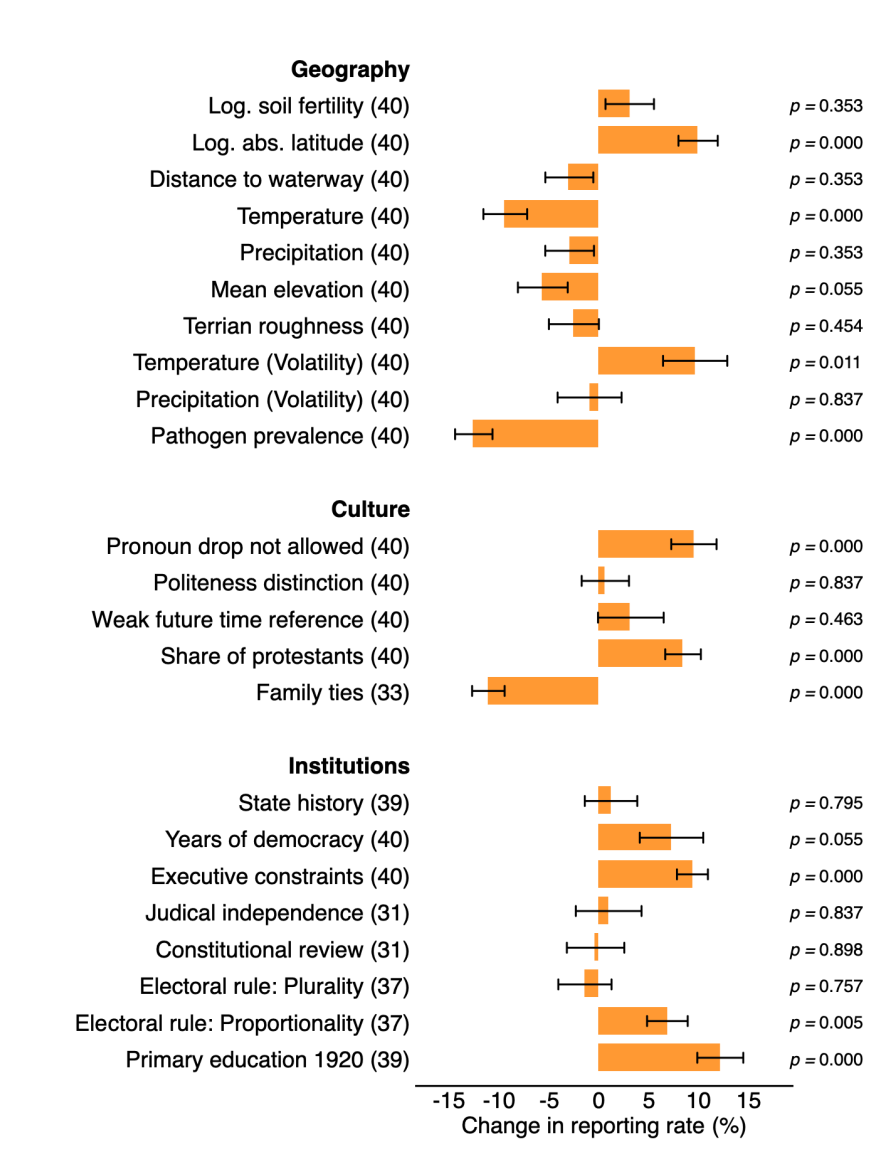


FIGURE A2.9
Correlates of Civic Honesty: Money

Notes. OLS coefficient estimates with standard errors clustered at the country-level. The sample is restricted to drop-offs in treatment Money. The dependent variable takes on the value 100 when an individual reported a wallet and 0 otherwise. Each coefficient has been estimated separately using standardized explanatory variables. They can therefore be interpreted as the difference in reporting rates associated with a one standard deviation change in the explanatory variable. We control for treatment status, institution fixed effects, and our standard set of control variables for recipient and situational characteristics: dummies for age above 40 years and gender, as well as the presence of a computer, coworkers, and other bystanders. To correct for multiple hypothesis testing, P -values are adjusted for false discovery rate (Benjamini and Hochberg, 1995). The number of countries included in the regressions is indicated in parentheses.

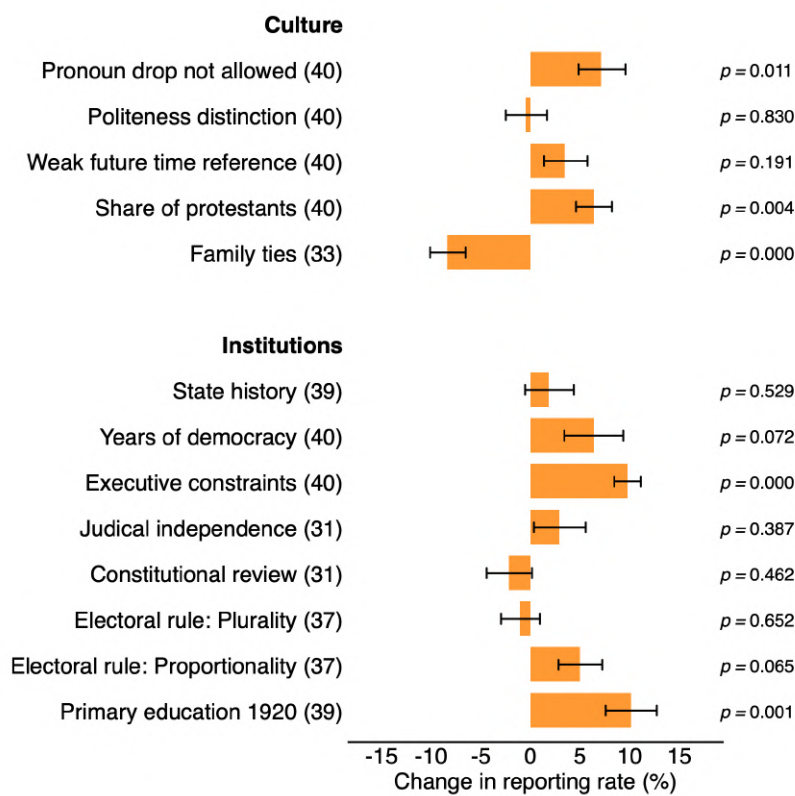


FIGURE A2.10

Correlates of Civic Honesty: Controlling for Geography

Notes. OLS coefficient estimates with standard errors clustered at the country-level. The dependent variable takes on the value 100 when an individual reported a wallet and 0 otherwise. Each coefficient has been estimated separately using standardized explanatory variables. They can therefore be interpreted as the difference in reporting rates associated with a one standard deviation change in the explanatory variable. We control for the first principal component of all geographical measures, treatment status, institution fixed effects, and our standard set of control variables for recipient and situational characteristics: dummies for age above 40 years and gender, as well as the presence of a computer, coworkers, and other bystanders. To correct for multiple hypothesis testing, P -values are adjusted for false discovery rate (Benjamini and Hochberg, 1995). The number of countries included in the regressions is indicated in parentheses.

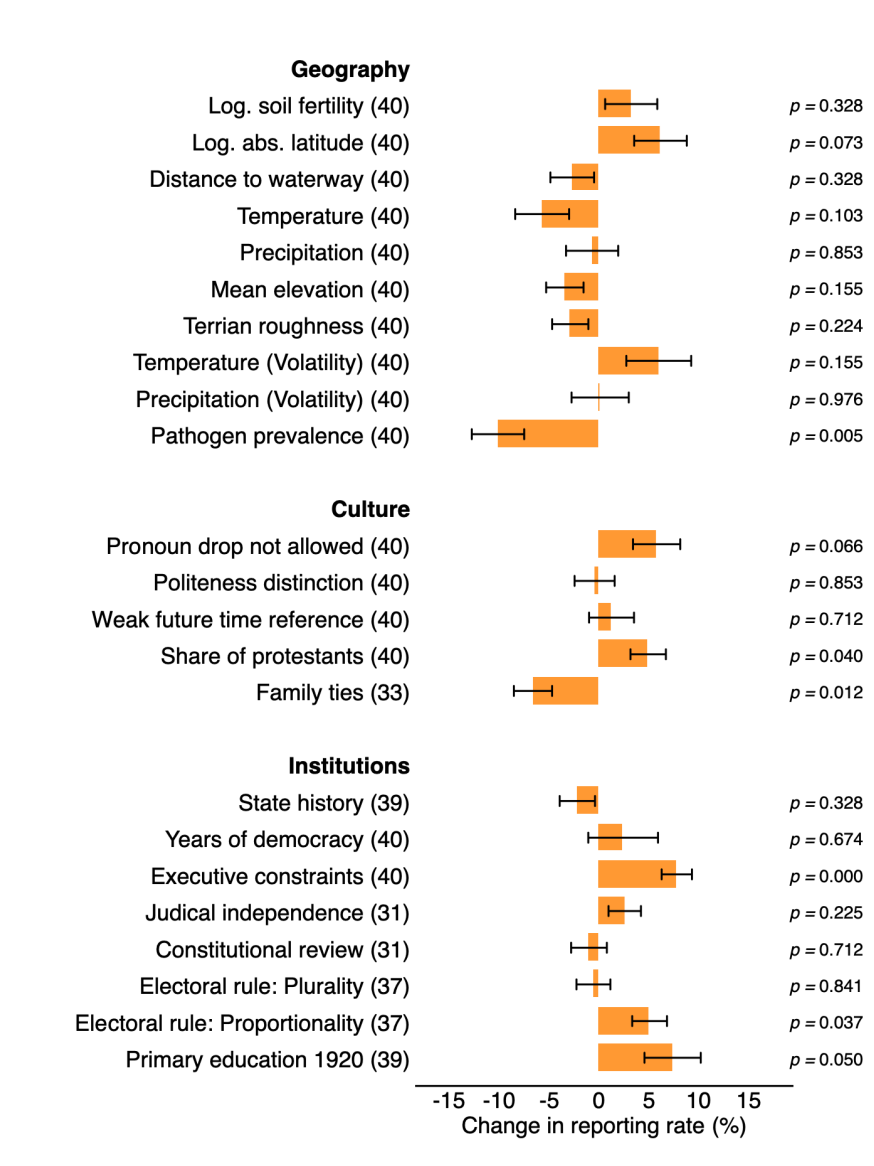


FIGURE A2.11

Correlates of Civic Honesty: Controlling for Country GDP

Notes. OLS coefficient estimates with standard errors clustered at the country-level. The dependent variable takes on the value 100 when an individual reported a wallet and 0 otherwise. Each coefficient has been estimated separately using standardized explanatory variables. They can therefore be interpreted as the difference in reporting rates associated with a one standard deviation change in the explanatory variable. We control for the logarithm of a country's GDP per capita in 2010 (IMF World Economic Outlook; based on purchasing-power-parity), treatment status, institution fixed effects, and our standard set of control variables for recipient and situational characteristics: dummies for age above 40 years and gender, as well as the presence of a computer, coworkers, and other bystanders. To correct for multiple hypothesis testing, *P*-values are adjusted for false discovery rate (Benjamini and Hochberg, 1995). The number of countries included in the regressions is indicated in parentheses.

IV. ALTERNATIVE EXPLANATIONS

We explored several alternative explanations for why rates of civic honesty tend to increase with greater amounts of money left in a wallet.

A. *Fear of Punishment*

One possibility is that wallet recipients were concerned about possible punishment for not reporting the wallet, especially when a wallet contained relatively more money. We purposefully designed our experiment to minimize such concerns by telling recipients that the wallet was found on a different street and having our research assistants immediately leave upon handing over the wallet (thereby never receiving written confirmation for the lost item). We also note that lost property laws tend to be uncommon and even when in place are rarely enforced (West, 2003).¹⁹

We first address the issue of punishment concerns by exploiting regional variation in lost property laws within the US. The US legal system is based on common law, under which a person who finds lost property can keep the item until the original owner comes forward.²⁰ However, some states have enacted statutes that modify the common law's treatment of lost property. For instance, the state of New York imposes a fine of up to one hundred dollars if a finder willfully fails to report lost property.²¹

About half of our lost wallet observations in the US originate from states that have adopted statutes explicitly requiring finders to return lost property to the rightful owner or to a relevant agency, such as the police. We therefore divided our sample according to whether legal consequences could ensue for failing to return a lost wallet. If fear of legal punishment drives the

19. In our representative survey we find a small but significant increase in self-reported fear of punishment with greater amounts of money in the wallet ($t_{2150} = 3.19$, $P = 0.001$, for the difference between the NoMoney and the Money condition; $t_{2150} = 2.45$, $P = 0.014$, for the difference between the Money and BigMoney condition). However, Column (4) in Table A2.11 shows that while theft aversion concerns and altruism are positively correlated with the intention to report the wallet, self-reported fear of punishment does not predict the stated likelihood of reporting the wallet.

20. Legal Information Institute, https://www.law.cornell.edu/wex/lost_property, accessed on September 18, 2016. Common law distinguishes between lost and mislaid property. Lost property is property that was unintentionally left behind by its owner. Mislaid property, on the other hand, is property that was intentionally set down in a location by its owner and then forgotten.

21. See N.Y. Personal Property Law § 252 (3).

TABLE A2.14
CIVIC HONESTY AND LOST PROPERTY LAWS

	Lost property law?	
	No (1)	Yes (2)
Money	16.809** (5.501)	20.350*** (5.657)
BigMoney	29.887*** (5.876)	25.576*** (6.245)
Constant	39.166** (12.925)	35.011** (11.245)
Controls:		
Recipient	yes	yes
Situation	yes	yes
Institution FE	yes	yes
City FE	yes	yes
Other treatments	yes	yes
Money = BigMoney	0.026	0.404
Observations	496	504
Adjusted R^2	0.152	0.055

Notes. OLS estimates with robust standard errors in parentheses. Column (1) focuses on US states without a lost property law, whereas Column (2) contains data from states with such a law. The dependent variable in both columns takes on the value 100 if a wallet was reported and 0 otherwise. “Money” and “BigMoney” are treatment indicators (we also include an indicator for treatment “Money-NoKey” but do not report its estimates for ease of exposition). Both models include binary control variables for individual and situational factors, including a recipient’s age (above 40 years) and gender (male), as well as the presence of a computer, coworkers, and other bystanders. The models also include city and institution fixed effects. The bottom of the table reports P -values from t -tests for equality of the treatment coefficients. Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

increase in reporting rates, then this relationship should be especially pronounced for states with lost property regulations. As shown in Table A2.14 however, we find similar treatment effects regardless of whether a state has a lost property law. Using seemingly unrelated regressions for states with and without property laws (Zellner, 1962), we fail to find a reliable difference in the size of the coefficients between the two groups for either the Money treatment ($\chi_1^2 = 0.21$, $P = 0.646$) or the BigMoney treatment ($\chi_1^2 = 0.27$, $P = 0.607$). Thus, recipients in states with legal sanctions surrounding lost property did not act in a meaningfully different way from recipients in states without such laws.

A second way we address possible punishment concerns is by examining whether the presence of a security camera moderates our results. Security cameras could serve as proof that the wallet was turned in to the recipient and therefore amplify concerns about punishment if the wallet was not returned. After each drop-off, except in Poland and the United Kingdom, our research assistants took note of whether they observed a security camera. Column (1) in Ta-

TABLE A2.15
CIVIC HONESTY AND PRESENCE OF SECURITY CAMERAS

	Full sample	Security camera?	
		No	Yes
	(1)	(2)	(3)
Money	10.558*** (0.732)	8.963*** (1.167)	11.591*** (0.950)
Security Camera	-2.659** (0.956)		
Constant	40.096*** (5.143)	27.774* (11.485)	38.699*** (5.691)
Recipient	Yes	Yes	Yes
Situation	Yes	Yes	Yes
Institution FE	Yes	Yes	Yes
City FE	Yes	Yes	Yes
Other treatments	Yes	Yes	Yes
Observations	15369	5806	9563
Adjusted R^2	0.189	0.224	0.170

Notes. OLS estimates with robust standard errors in parentheses. Column (1) shows the estimates for the full sample as a benchmark, Column (2) contains observations where no security camera was sighted, and Column (3) includes only observations where a camera was sighted. The dependent variable in all models takes on the value 100 if a wallet was reported and 0 otherwise. “Money” is a dummy for treatment Money (we also include indicators for treatments “Money-NoKey” and “BigMoney” but do not report their estimates for ease of exposition). All models include binary control variables for recipient and situational characteristics, including a recipient’s age (above 40 years), gender (male), and the presence of a computer, coworkers and other bystanders. The models also include city and institution fixed effects. Note that the sample does not include data from the United Kingdom and Poland because we did not collect data on security cameras. Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

ble A2.15 shows that if anything, the presence of a security camera during the drop-off lowered the likelihood of reporting a wallet by 2.7 percentage points ($t_{15044} = -2.78$, $P = 0.005$). While the treatment effect in the Money condition, relative to the NoMoney condition, is slightly larger for drop-off locations with cameras than those without ($\chi_1^2 = 3.20$, $P = 0.074$ when comparing the Money coefficient in Columns 2 and 3), the treatment effect is large and significant for both subsamples ($t_{5485} = 7.68$, $P < 0.001$ for Column 2; $t_{9241} = 12.20$, $P < 0.001$ for Column 3).

A third approach we use to address punishment concerns involves the presence of other individuals when performing a wallet drop-off. Recipients may have been worried about negative reactions from bystanders—an informal punishment—for not reporting a wallet. After performing the wallet drop-offs, our research assistants also noted whether coworkers and other individuals were present during the exchange. If worries about informal sanctions influenced recipient’s behavior then we should observe smaller treatment effects when other individuals were not present.

TABLE A2.16
CIVIC HONESTY AND SOCIAL MONITORING

	Full sample	No coworkers	No bystanders	Alone
	(1)	(2)	(3)	(4)
Money	10.792*** (0.712)	9.944*** (0.884)	10.083*** (1.216)	8.824*** (1.506)
Constant	33.302** (11.112)	28.147* (11.407)	64.166** (24.972)	67.158** (24.791)
Controls:				
Recipient	yes	yes	yes	yes
Situation	yes	yes	yes	yes
Institution FE	yes	yes	yes	yes
City FE	yes	yes	yes	yes
Other treatments	yes	yes	yes	yes
Observations	17295	11528	5939	4079
Adjusted R^2	0.185	0.178	0.205	0.196

Notes. OLS estimates with robust standard errors in parentheses. Column (1) shows the estimates for the full sample as a benchmark, Column (2) includes observations without coworkers present, Column (3) includes observations without other bystanders present, and Column (4) includes observations where neither coworkers nor other bystanders were present. The dependent variable in all models takes on the value 100 if a wallet was reported and 0 otherwise. “Money” is a dummy for treatment Money (we also include an indicators for treatments “Money-NoKey” and “BigMoney” but do not report their estimates for ease of exposition). All models include binary control variables for recipient and situational characteristics, including a recipient’s age (above 40 years), gender (male), and the presence of a computer. The models also include city and institution fixed effects. Significance levels: * $P < 0.05$, ** $P \leq 0.01$, *** $P < 0.001$.

Table A2.16 displays the results for the full sample compared to instances when no coworkers were present, no bystanders were present, and when the recipient and research assistant were completely alone during the exchange. Relative to the full sample, we fail to find a reliable difference in treatment effects when co-workers are not present ($\chi_1^2 = 0.56$, $P = 0.455$ when comparing the coefficient of Money in Columns 1 and 2), when other individuals are not present ($\chi_1^2 = 0.25$, $P = 0.615$ comparing Columns 1 and 3), and when recipients were alone ($\chi_1^2 = 1.40$, $P = 0.238$ comparing Columns 1 and 4). We observe roughly similarly-sized treatment effects between the full sample and all subsamples, suggesting that the presence of others did not qualify our results.

B. Returning the Wallet but Pocketing the Money

Another explanation for our main result is that recipients in the Money and BigMoney conditions may have been more likely to return the wallet after first pocketing the money. We decided not to collect reported wallets to minimize the inconvenience to the recipients. It is

possible that some recipients contacted the owner to return the wallet without the money.

To examine this possibility we picked up all reported wallets in seven cities across the Czech Republic (82 wallets) and Switzerland (90 wallets). We selected these two countries because they differ markedly in their level of corruption and presumably also in dishonest behavior.²² If some recipients reported the wallet after first pocketing the money, then we should observe wallets that are returned without any money (especially in the Czech Republic where corruption is more prevalent). However, we recovered 99% and 98% of the money from the wallets that we picked up in Switzerland and the Czech Republic, respectively, and we observe no reliable difference between the two countries ($Z = 0.22$, $P = 0.823$ by a rank-sum test). This suggests that collecting emails was a valid method to measure whether people would return a wallet with all of its contents.

C. Possible Finder's Fee for Returning a Wallet

Another explanation for the increase in civic honesty for wallets with greater amounts of money is that the recipients expected a larger monetary reward (i.e., “finder’s fee”) when returning a wallet that contained relatively more money. To examine this possibility, we asked respondents in our representative survey experiments about their beliefs regarding a finder’s fee and find results that are inconsistent with the behavioral patterns from our field experiments.

In the representative survey experiments, we asked the participants to estimate the likelihood that they would receive a financial reward from the owner, and if they received such a reward, how much money did they think they would get. We constructed a measure of *expected reward* by multiplying these two estimates together, and to facilitate comparability across countries we converted amounts to US dollars using the same exchange rate as in our field experiments. Overall, 42% of the participants stated that they would not expect a financial reward at all. The median expected reward ranged between US \$0.00 (Money-NoKey condition) and \$1.58 (High-Stakes condition)—cash amounts that were much lower than what the recipients could have gained from keeping the wallet (except for the NoMoney condition). Finally, we do

22. In 2013, Transparency International ranked Switzerland 7th and the Czech Republic 57th out of 177 countries.

TABLE A2.17
CIVIC HONESTY AND BELIEFS ABOUT FINDER'S FEES

	Expected reward (in US \$)	Reporting likelihood (in %)
	(1)	(2)
Money	-1.833*** (0.511)	2.417* (0.998)
BigMoney	-0.128 (0.460)	3.848*** (0.976)
Money-NoKey	-2.816*** (0.397)	-2.428* (1.184)
Expected reward (in US \$)		0.009 (0.039)
Constant	3.995*** (0.516)	86.378*** (1.249)
Controls:		
Institution FE	yes	yes
Country FE	yes	yes
Money = BigMoney	0.000	0.125
Money = Money-NoKey	0.008	0.000
Observations	2160	2160
Adjusted R^2	0.028	0.029
F	13.320	6.619

Notes. OLS estimates with robust standard errors in parentheses. In Column (1), the dependent variable is participants' expected financial reward for reporting the wallet (in US dollars). The dependent variable in Column (2) is the likelihood that participants would report the wallet (as a percentage). "Money," "BigMoney," and "Money-NoKey," are treatment indicators. All models include country and institution fixed effects. The bottom of the table reports P -values from t -tests for equality of the treatment coefficients. Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

not observe that the expected reward increased monotonically with the amount of money in the wallet, as shown in Column (1) of Table A2.17. In fact, on average participants expected the highest reward in the NoMoney condition.²³ We also do not find that a higher expected reward is associated with a higher stated likelihood of reporting the wallet, as shown in Column (2) of Table A2.17. Moreover, controlling for a respondent's expected reward does not meaningfully change our observed treatment effects (see Column 2 of Table A2.11 for comparison). Overall, the prospect of a financial reward is unlikely to explain the monotonic increase in reporting rates.

23. One potential explanation for this seemingly counterintuitive result is that the amount of cash in the Money condition serves as an upper bound on the amount people expect to receive as a finder's fee. Consistent with this interpretation, when examining conditional expectations about the reward (i.e., how much money a respondent expects to receive, conditional upon receiving a reward for returning the wallet) we find that only 12% of responses exceeded \$13.45 in the Money condition, compared to 35% of responses in the NoMoney condition ($Z = 9.13$, $P < 0.001$). This difference is significant when examining each country (US, UK, and Poland) separately ($Z > 4.00$ in all conditions, $P < 0.001$). Furthermore, and also consistent with a censoring effect, we observed greater variability in conditional expected finder fees in the NoMoney condition than in any other condition ($P < 0.001$ by a variance-ratio test for every pairwise comparison between the NoMoney condition and all other conditions).

V. ROBUSTNESS CHECKS

A. *Individual and Situational Factors*

To what extent do individual and situational factors drive cross-country differences in civic honesty? For instance, drop-off locations may have been more crowded in some countries with the possible consequence that recipients felt more observed and obliged to return the wallet. Or perhaps recipients were busier when there were more customers present during the drop-off and as a result less likely to report a wallet.

To examine the robustness of cross-country differences in civic honesty, we estimated the residuals from a regression that accounted for recipient and situational characteristics between locations as well as institution fixed effects. We conducted this analysis separately for the Money and NoMoney conditions, and then aggregated the residuals by country. For ease of exposition, we add the average reporting rate across all countries. The resulting regression-adjusted ranking and the original country ranking were virtually the same for both the NoMoney and Money conditions (Spearman's $\rho = 0.976$ and 0.990 , respectively; both P -values < 0.001). Moreover, the range of reporting rates across countries remained large and almost identical when using the regression-adjusted data instead of the original data (Figure A2.12). This suggests that differences in recipient and situational characteristics between locations did not account for large differences in civic honesty across countries.

B. *Experimenter Effects*

We also examined the role of our research assistants in influencing recipient decisions to report the wallets. We used a total of 13 research assistants (all recruited from two German speaking universities), and purposely created overlaps for some of the countries they traveled to. We had two research assistants with overlapping presences in France, Germany, Italy, Malaysia, Poland, Spain, Switzerland, Turkey and the UK, and seven research assistants in the US. Table A2.18 presents an overview of the number of wallets each research assistant turned in by country. The numbers in parentheses represent the number of wallets for which there was at

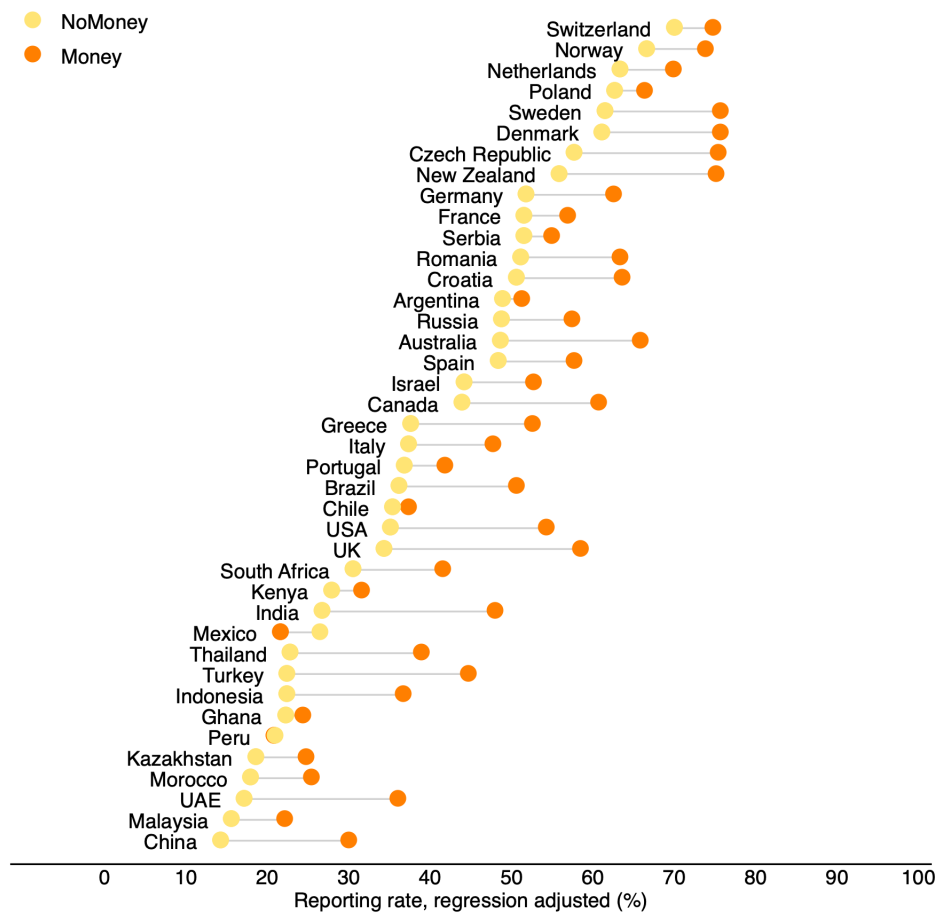


FIGURE A2.12
Regression-Adjusted Ranking

Notes. Regression-adjusted share of wallets reported in the NoMoney (US \$0) and Money (US \$13.45) condition by country. We regress individual decisions to report a wallet on recipient (age and gender of the recipient) and situational control variables (presence of a computer, number of coworkers and other bystanders) as well as institution fixed effects, and subsequently computed residuals for treatment Money and NoMoney. Finally, we aggregated residuals for each country and added the overall average reporting rate. The original and the regression-adjusted ranking are highly correlated for both the NoMoney and Money conditions (Spearman's $\rho = 0.976$, $P < 0.001$ and $\rho = 0.990$, $P < 0.001$, respectively).

least one other research assistant performing drop-offs in the same city. These overlaps help us to distinguish between experimenter and city fixed effects.

We first explored the influence of research assistants by introducing experimenter fixed effects in our benchmark regression model. Tables A2.19 and A2.20 present the estimates of the treatment effects with and without experimenter fixed effects for each country where we had an overlap. We ran several tests to assess the influence of the research assistants. First, we found

that the treatment effects in each country remained basically the same, regardless of whether we control for experimenter fixed effects.²⁴ Second, in the US we performed all 21 pairwise comparisons of the seven experimenter fixed effects and found that none of the comparisons are statistically significant at the 5% level (note that this is a conservative test since we do not adjust the P -values for multiple hypothesis testing). Third, we conducted joint significance tests of the experimenter fixed effects and found null results in all countries (F -tests in Tables A2.19 and A2.20). Finally, we computed the change in the variance explained (measured by the adjusted R^2) when we augment our benchmark specification with experimenter fixed effects and found virtually no change in the variance explained (as shown at the bottom of Tables A2.19 and A2.20). Overall, we find little evidence that differences between research assistants are driving our results.

24. We also estimated the same regression model as in Column 2 of Table A2.8 and added the experimenters' age and gender as explanatory variables. Both coefficients failed to reach statistical significance, suggesting that experimenter age and gender did not reliably influence reporting rates among recipients ($t_{16924} = 1.17$, $P = 0.243$ for age; $t_{16924} = 1.18$, $P = 0.237$ for gender). We also failed to find a significant interaction effect between the gender of the experimenter and gender of the recipient ($t_{16923} = 0.90$, $P = 0.368$). However, these null results for experimenter gender should be interpreted carefully given that we only had two female research assistants.

TABLE A2.18
DROP-OFFS BY EXPERIMENTERS AND COUNTRY: OVERLAPS

Experimenter	Country									
	France	Germany	Italy	Malaysia	Poland	Spain	Switzerland	Turkey	United Kingdom	United States
#1		75 (75)	202 (50)				65 (65)	64 (64)		90 (90)
#2		325 (76)				71 (53)				
#3	385 (111)						334 (63)			
#4						329 (57)			360 (107)	
#5										62 (62)
#6										75 (75)
#7										
#8							109 (109) 291 (109)			86 (86)
#9			198 (46)		399 (42) 401 (42)				772 (96)	
#10	417 (111)									
#11								336 (118)		569 (74)
#12										58 (58)
#13										60 (60)
Total obs.	802	400	400	400	800	400	399	400	1132	1000

Notes. Number of wallets turned in by experimenter and country in countries with overlaps. The number of drop-offs where at least one other experimenter turned in wallets in the same city is in parenthesis.

TABLE A2.19
EXPERIMENTER EFFECTS

	France		Germany		Italy		Malaysia		Poland	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Money	5.423 (3.394)	5.449 (3.395)	10.651* (4.385)	10.629* (4.388)	11.683* (4.747)	11.567* (4.771)	5.791 (3.884)	5.758 (3.908)	3.310 (4.690)	3.311 (4.692)
BigMoney									11.761** (4.410)	11.790** (4.422)
Constant	63.758*** (11.491)	63.393*** (11.508)	77.793*** (9.841)	82.424*** (10.131)	49.231*** (14.620)	45.81251* (17.275)	41.396*** (10.701)	40.889*** (11.401)	59.380*** (11.216)	58.974*** (11.198)
Controls:										
Experimenter FE	no	yes	no	yes	no	yes	no	yes	no	yes
Recipient	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Situation	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Institution FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
City FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Other treatment	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
F-test Experimenter effects		0.537		0.185		0.700		0.894		0.202
Money = BigMoney									0.058	0.058
Observations	802	802	400	400	400	400	400	400	794	794
Adjusted R^2	0.074	0.073	0.185	0.186	0.091	0.089	0.050	0.048	0.050	0.051

Notes. OLS estimates with robust standard errors in parentheses. The dependent variable in all models takes on the value 100 if a wallet was reported and 0 otherwise. “Money,” “BigMoney,” and “Money-NoKey” are treatment indicators. All models include binary control variables for individual and situational factors, including a recipient’s age (above 40 years) and gender (male), the presence of a computer, other people, and coworkers. All models include city and institution fixed effects. Models in the even columns additionally control for experimenter fixed effects. The bottom of the table reports P -values from t -tests for equality of the coefficients of treatments “Money” and “BigMoney.” Significance levels: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A2.20
EXPERIMENTER EFFECTS (CONTINUED)

	Spain		Switzerland		Turkey		United Kingdom		United States	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Money	9.479 (4.899)	10.086* (4.920)	4.094 (4.178)	4.194 (4.187)	20.087*** (4.486)	20.160*** (4.492)	23.106*** (3.851)	23.074*** (3.851)	18.301*** (3.934)	18.235*** (3.955)
BigMoney							35.941*** (4.567)	35.929*** (4.571)	27.832*** (4.260)	27.793*** (4.247)
Constant	30.454* (15.482)	20.523 (17.847)	62.645*** (11.165)	66.160*** (12.697)	12.633 (11.706)	10.883 (14.075)	25.763** (9.345)	28.239* (11.243)	34.445** (11.291)	39.464** (13.585)
Controls:										
Experimenter FE	no	yes	no	yes	no	yes	no	yes	no	yes
Recipient	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Situation	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Institution FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
City FE	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
Other treatment	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
F-test Experimenter effects		0.271		0.579		0.819		0.699		0.587
Money = BigMoney							0.001	0.001	0.027	0.027
Observations	400	400	399	399	400	400	1132	1132	1000	1000
Adjusted R^2	0.046	0.047	0.056	0.054	0.086	0.083	0.122	0.122	0.100	0.099

Notes. OLS estimates with robust standard errors in parentheses. The dependent variable in all models takes on the value 100 if a wallet was reported and 0 otherwise. “Money,” “BigMoney,” and “Money-NoKey” are treatment indicators. All models include binary control variables for individual and situational factors, including a recipient’s age (above 40 years) and gender (male), the presence of a computer, other people, and coworkers. All models include city and institution fixed effects. Models in the even columns additionally control for experimenter fixed effects. The bottom of the table reports P -values from t -tests for equality of the coefficients of treatments “Money” and “BigMoney.” Significance levels: * $P \leq 0.05$, ** $P < 0.01$, *** $P < 0.001$.

C. Differences in Email Usage

Since our measure of civic honesty relied on recipients contacting the owner by email, one concern is that differences in exposure to email communication could be responsible for cross-country differences in reporting rates. Yet, we focused on drop-off locations in urban places and included institutions where email communication is common. In particular, hotel staff should be able to communicate via email in all parts of the world. Consequently, if email experience is a key driver of differences in reporting rates, we should see substantially less heterogeneity when we restrict our sample to hotels. However, Figure A2.13 shows that this is not the case. We still observe large differences in reporting rates across countries when focusing on hotels only. As a further robustness check, we included the share of firms that use email to interact with their customers and suppliers in a country (from the World Bank Global Enterprise Survey) as an additional control variable to construct the regression-adjusted measure of civic honesty.²⁵ Figure A2.14 shows that the differences between countries remain large, and the regression-adjusted ranking is almost identical to the unconditional ranking (Spearman's $\rho = 0.950$, $P < 0.001$ for the NoMoney condition, and $\rho = 0.932$, $P < 0.001$ for the Money condition). This suggests that experience with email communication was not a major driver of cross-country differences in reporting rates.

D. Differences in Economic Development

We also assessed the extent that cross-country variation in civic honesty was robust when controlling for differences in economic development. For this purpose, we included contemporary per capita income as an additional control variable for the estimation of regression-adjusted reporting rates. The results in Figure A2.15 demonstrate that cross-country differences remain substantial, even when controlling for economic development. The regression-adjusted rankings from Figure A2.15 are also positively correlated with the unconditional rankings from 2.1 (Spearman's $\rho = 0.705$, $P < 0.001$ for the NoMoney condition; Spearman's $\rho = 0.753$, $P < 0.001$ for the Money condition).

25. The Global Enterprise Survey does not cover most Western European countries and North America, so we limit our analysis of email usage to 27 countries.

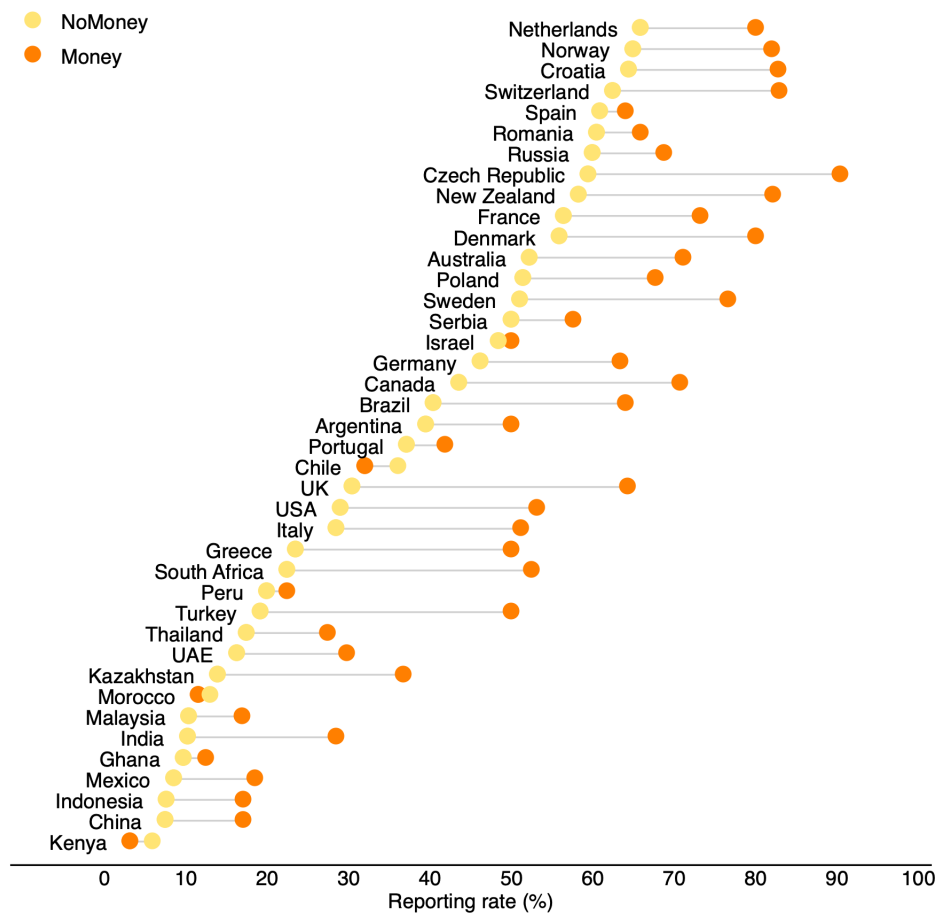


FIGURE A2.13
Country Ranking for Hotels

Notes. Share of wallets reported by hotel employees in treatments NoMoney (US \$0) and Money (US \$13.45) by country. The amount of money in the wallet is adjusted to purchasing power parity for each country. 'AVERAGE' shows the averages across all 40 countries.

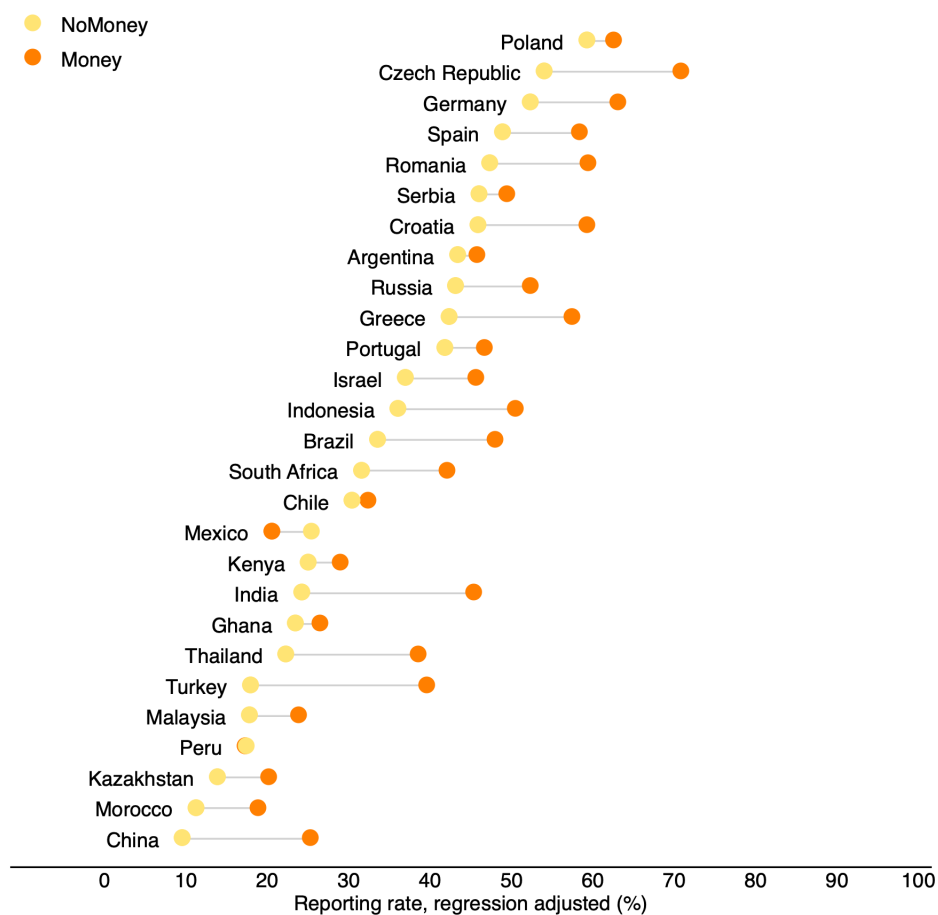


FIGURE A2.14
Regression-Adjusted Ranking: Email Usage

Notes. Regression-adjusted share of wallets reported in treatment decisions to report a wallet on the share of firms that use email to interact with their customers and suppliers in a country (from the World Bank Global Enterprise Survey), individual (age and gender of the recipient) and situational control variables (presence of a computer, number of coworkers and other bystanders) as well as institution fixed effects, and subsequently computed residuals for treatments Money and NoMoney. Finally, we aggregated residuals for each country and added the overall average reporting rate. The regression-adjusted ranking is almost identical to the unconditional ranking (Spearman's $\rho = 0.950$, $P < 0.001$ for treatment NoMoney, and $\rho = 0.932$, $P < 0.001$ for treatment Money). Due to missing data, the estimates are based on a sample of 27 countries.

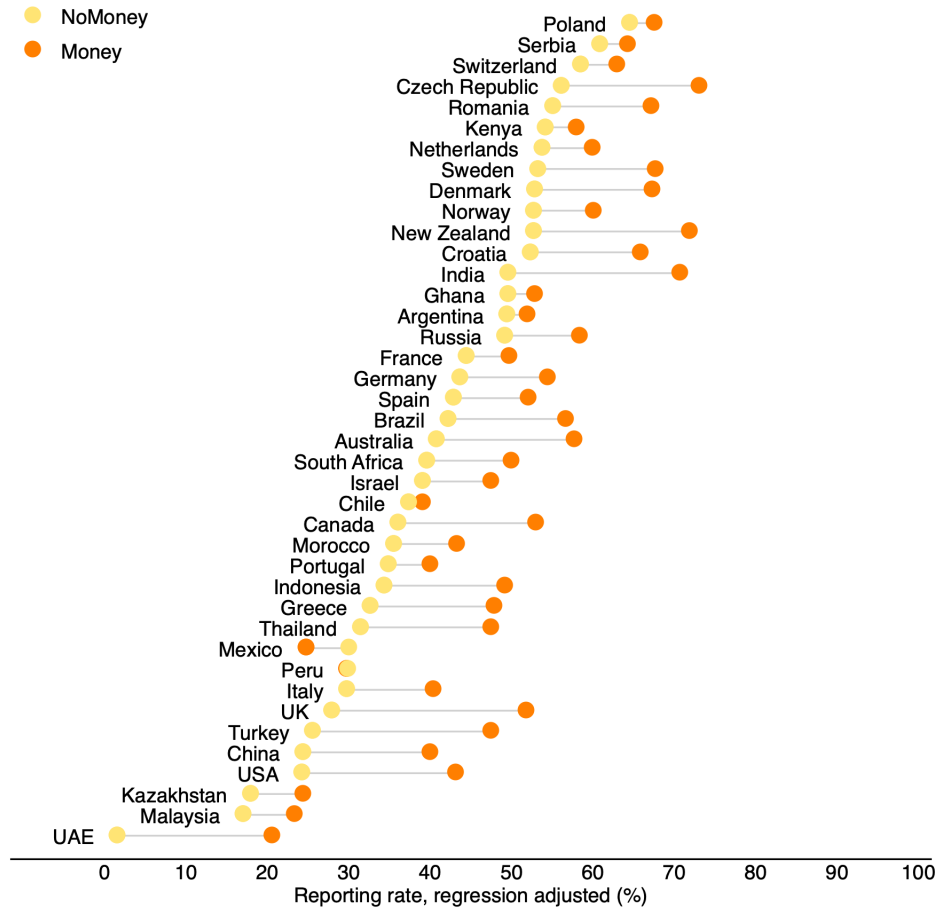


FIGURE A2.15
Regression-Adjusted Ranking: Country GDP

Notes. Regression-adjusted share of wallets reported in the NoMoney (US \$0) and Money (US \$13.45) condition by country. We regress individual decisions to report a wallet on the logarithm of a country's GDP per capita in 2010 (IMF World Economic Outlook; based on purchasing-power-parity), in addition to recipient (age and gender of the recipient) and situational control variables (presence of a computer, number of coworkers and other bystanders) as well as institution fixed effects. We subsequently computed residuals for treatment Money and NoMoney. Finally, we aggregated residuals for each country and added the overall average reporting rate. The original and the GDP-adjusted ranking are significantly correlated for both the NoMoney and Money conditions (Spearman's $\rho = 0.705$, $P < 0.001$ and $\rho = 0.753$, $P < 0.001$, respectively).

REFERENCES

- ACEMOGLU, D., S. JOHNSON, AND J. A. ROBINSON (2002): "Reversal of fortune: Geography and institutions in the making of the modern world income distribution," *Quarterly Journal of Economics*, 117, 1231–1294.
- AGHION, P., Y. ALGAN, P. CAHUC, AND A. SHLEIFER (2010): "Regulation and distrust," *Quarterly Journal of Economics*, 125, 1015–1049.
- ALESINA, A. AND P. GIULIANO (2010): "The power of the family," *Journal of Economic Growth*, 15, 93–125.
- (2014): "Family ties," in *Handbook of Economic Growth*, Elsevier, vol. 2, 177–215.
- (2015): "Culture and institutions," *Journal of Economic Literature*, 53, 898–944.
- ALESINA, A., P. GIULIANO, AND N. NUNN (2013): "On the origins of gender roles: Women and the plough," *Quarterly Journal of Economics*, 128, 469–530.
- ALGAN, Y. AND P. CAHUC (2013): "Trust and growth," *Annual Review of Economics*, 5, 521–549.
- ANANYEV, M. AND S. GURIEV (forthcomming): "Effect of income on trust: Evidence from the 2009 crisis in Russia," *The Economic Journal*.
- ANDREONI, J. AND B. D. BERNHEIM (2009): "Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects," *Econometrica*, 77, 1607–1636.
- ANDREONI, J. AND J. MILLER (2002): "Giving according to GARP: An experimental test of the consistency of preferences for altruism," *Econometrica*, 70, 737–753.
- ARRUÑADA, B. (2010): "Protestants and Catholics: Similar work ethic, different social ethic," *Economic Journal*, 120, 890–918.
- ASHRAF, Q. AND O. GALOR (2011a): "Cultural diversity, geographical isolation, and the origin of the wealth of nations," *National Bureau of Economic Research*, 17640.
- (2011b): "Dynamics and stagnation in the Malthusian epoch," *American Economic Review*, 101, 2003–41.
- (2013): "The 'Out of Africa' hypothesis, human genetic diversity, and comparative economic development'," *American Economic Review*, 103, 1–46.
- BANFIELD, E. C. (1958): *The Moral Basis of a Backward Society*, Free Press.
- BATTIGALLI, P. AND M. DUFWENBERG (2007): "Guilt in games," *American Economic Review*, 97, 170–176.
- BECK, T., G. CLARKE, A. GROFF, P. KEEFER, AND P. WALSH (2001): "New tools in comparative political economy: The database of political institutions," *World Bank Economic Review*, 15, 165–176.
- BÉNABOU, R. AND J. TIROLE (2006): "Incentives and prosocial behavior," *American Economic Review*, 96, 1652–1678.
- (2011): "Identity, morals, and taboos: beliefs as assets," *Quarterly Journal of Economics*, 126, 805–855.

- BENAVOT, A. AND P. RIDDLE (1988): "The expansion of primary education, 1870-1940: Trends and issues," *Sociology of Education*, 61, 191–210.
- BENJAMINI, Y. AND Y. HOCHBERG (1995): "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B*, 57, 289–300.
- BISIN, A. AND T. VERDIER (2011): "The economics of cultural transmission and socialization," in *Handbook of Social Economics*, Elsevier, vol. 1, 339–416.
- BOCKSTETTE, V., A. CHANDA, AND L. PUTTERMAN (2002): "States and markets: The advantage of an early start," *Journal of Economic Growth*, 7, 347–369.
- BROWN, R. AND A. GILMAN (1960): "The pronouns of power and solidarity," in *Style in Language*, ed. by T. A. Sebeok, MIT Press, 253–276.
- BUGGLE, J. C. AND R. DURANTE (2017): "Climate Risk, Cooperation, and the Co-Evolution of Culture and Institutions," *CEPR Discussion Paper*, 12380.
- CASHDAN, E. A. (1985): "Coping with risk: Reciprocity among the Basarwa of Northern Botswana," *Man*, 20, 454–474.
- CASSAR, A., G. D'ADDA, AND P. GROSJEAN (2014): "Institutional quality, culture, and norms of cooperation: Evidence from behavioral field experiments," *Journal of Law and Economics*, 57, 821–863.
- CHARNESS, G. AND M. DUFWENBERG (2006): "Promises and partnership," *Econometrica*, 74, 1579–1601.
- CHARNESS, G. AND M. RABIN (2002): "Understanding social preferences with simple tests," *Quarterly Journal of Economics*, 117, 817–869.
- CHEN, M. K. (2013): "The effect of language on economic behavior: Evidence from savings rates, health behaviors, and retirement assets," *American Economic Review*, 103, 690–731.
- COLEMAN, J. S. (1990): *Foundations of Social Capital Theory*, Cambridge, MA: Harvard University Press.
- DAVIS, M. H. (1983): "Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach," *Journal of Personality and Social Psychology*, 44, 113–126.
- DEAN, J. S., R. C. EULER, G. J. GUMERMAN, F. PLOG, R. H. HEVLY, AND T. N. KARLSTROM (1985): "Human behavior, demography, and paleoenvironment on the Colorado Plateaus," *American Antiquity*, 50, 537–554.
- DURANTE, R. (2010): "Risk, cooperation and the economic origins of social trust: an empirical investigation," .
- ENGEL, C. (2011): "Dictator games: a meta study," *Experimental Economics*, 14, 583–610.
- ENGERMAN, S. L. AND K. L. SOKOLOFF (1997): "Factor endowments, institutions, and differential paths of growth among new world economies," in *How Latin America Fell Behind: Essays on the Economic Histories of Brazil and Mexico, 1800-1914*, ed. by S. Haber, Stanford, CA: Stanford University Press, 260–304.
- ENKE, B. (forthcoming): "Kinship, cooperation, and the evolution of moral systems," *Quarterly Journal of Economics*.

- FALK, A., A. BECKER, T. DOHMEN, B. ENKE, D. HUFFMAN, AND U. SUNDE (2018): "Global evidence on economic preferences," *The Quarterly Journal of Economics*, 133, 1645–1692.
- FINCHER, C. L. AND R. THORNHILL (2012): "Parasite-stress promotes in-group assortative sociality: The cases of strong family ties and heightened religiosity," *Behavioral and Brain Sciences*, 35, 61–79.
- FINCHER, C. L., R. THORNHILL, D. R. MURRAY, AND M. SCHALLER (2008): "Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism," *Proceedings of the Royal Society of London B: Biological Sciences*, 275, 1279–1285.
- FISMAN, R., P. JAKIELA, AND S. KARIV (2015): "How did distributional preferences change during the great recession?" *Journal of Public Economics*, 128, 84–95.
- FUND, I. M. (2015): *World Economic Outlook, April 2015*, International Monetary Fund Washington, DC.
- GALLUP, J. L., J. D. SACHS, AND A. D. MELLINGER (1999): "Geography and economic development," *International Regional Science Review*, 22, 179–232.
- GALOR, O. AND Ö. ÖZAK (2016): "The agricultural origins of time preference," *American Economic Review*, 106, 3064–3103.
- GALOR, O. AND V. SAVITSKIY (2018): "Climatic roots of loss aversion," *CESifo Working Paper*.
- GLAESER, E. L., R. LA PORTA, F. LOPEZ-DE SILANES, AND A. SHLEIFER (2004): "Do institutions cause growth?" *Journal of Economic Growth*, 9, 271–303.
- GLAESER, E. L., D. I. LAIBSON, J. A. SCHEINKMAN, AND C. L. SOUTTER (2000): "Measuring trust," *Quarterly Journal of Economics*, 115, 811–846.
- GLAESER, E. L., G. A. PONZETTO, AND A. SHLEIFER (2007): "Why does democracy need education?" *Journal of Economic Growth*, 12, 77–99.
- GORODNICHENKO, Y. AND G. ROLAND (2017): "Culture, institutions, and the wealth of nations," *Review of Economics and Statistics*, 99, 402–416.
- GREEN, A. (1990): "Education and state formation." in *Education and State Formation*, Palgrave Macmillan, London, 76–110.
- GREIF, A. (1994): "Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies," *Journal of Political Economy*, 102, 912–950.
- (2006): *Institutions and the path to the modern economy: Lessons from medieval trade*, Cambridge University Press.
- GUIO, L., P. SAPIENZA, AND L. ZINGALES (2011): "Civic capital as the missing link," in *Handbook of Social Economics*, ed. by J. Benhabib, A. Bisin, and M. O. Jackson, Amsterdam: North-Holland, vol. 1, 417–480.
- (2016): "Long-term persistence," *Journal of the European Economic Association*, 14, 1401–1436.
- HELLIWELL, J. F. AND R. D. PUTNAM (2007): "Education and social capital," *Eastern Economic Journal*, 33, 1–19.
- HENRICH, J., R. BOYD, S. BOWLES, C. CAMERER, E. FEHR, H. GINTIS, AND R. MCELREATH (2001): "In search of Homo Economicus: behavioral experiments in 15 small-scale societies," *American Economic Review*, 91, 73–78.

- HENRICH, J., J. ENSMINGER, R. MCELREATH, A. BARR, C. BARRETT, A. BOLYANATZ, J. C. CARDENAS, M. GURVEN, E. GWAKO, N. HENRICH, ET AL. (2010): "Markets, religion, community size, and the evolution of fairness and punishment," *Science*, 327, 1480–1484.
- KASHIMA, E. S. AND Y. KASHIMA (1998): "Culture and language: The case of cultural dimensions and personal pronoun use," *Journal of Cross-Cultural Psychology*, 29, 461–486.
- KNACK, S. AND P. KEEFER (1997): "Does social capital have an economic payoff? A cross-country investigation," *Quarterly Journal of Economics*, 112, 1251–1288.
- KRUPKA, E. L. AND R. A. WEBER (2013): "Identifying social norms using coordination games: Why does dictator game sharing vary?" *Journal of the European Economic Association*, 11, 495–524.
- LA PORTA, R., F. LOPEZ-DE SILANES, C. POP-ELECHES, AND A. SHLEIFER (2004): "Judicial checks and balances," *Journal of Political Economy*, 112, 445–470.
- LA PORTA, R., F. LOPEZ-DE SILANES, A. SHLEIFER, AND R. VISHNY (1999): "The quality of government," *Journal of Law, Economics, and Organization*, 15, 222–279.
- LA PORTA, R., F. LOPEZ-DE SILANES, A. SHLEIFER, AND R. W. VISHNY (1997): "Trust in large organizations," *American Economic Review*, 87, 333.
- LICHT, A. N., C. GOLDSCHMIDT, AND S. H. SCHWARTZ (2007): "Culture rules: The foundations of the rule of law and other norms of governance," *Journal of Comparative Economics*, 35, 659–688.
- LITINA, A. (2016): "Natural land productivity, cooperation and comparative development," *Journal of Economic Growth*, 21, 351–408.
- LOWES, S., N. NUNN, J. A. ROBINSON, AND J. L. WEIGEL (2017): "The evolution of culture and institutions: Evidence from the Kuba kingdom," *Econometrica*, 85, 1065–1091.
- MICHALOPOULOS, S. (2012): "The origins of ethnolinguistic diversity," *American Economic Review*, 102, 1508–39.
- MILLIGAN, K., E. MORETTI, AND P. OREOPOULOS (2004): "Does education improve citizenship? Evidence from the United States and the United Kingdom," *Journal of Public Economics*, 88, 1667–1695.
- MOKYR, J. (2008): "The institutional origins of the industrial revolution," in *Institutions and Economic Performance*, ed. by E. Helpman, Harvard University Press, 64–119.
- MURRAY, D. R. AND M. SCHALLER (2010): "Historical prevalence of infectious diseases within 230 geopolitical regions: A tool for investigating origins of culture," *Journal of Cross-Cultural Psychology*, 41, 99–108.
- NANNICINI, T., A. STELLA, G. TABELLINI, AND U. TROIANO (2013): "Social capital and political accountability," *American Economic Journal: Economic Policy*, 5, 222–50.
- NETTLE, D. (1998): "Explaining global patterns of language diversity," *Journal of anthropological archaeology*, 17, 354–374.
- NORDHAUS, W. D. (2006): "Geography and macroeconomics: New data and new findings," *Proceedings of the National Academy of Sciences*, 103, 3510–3517.
- NUNN, N. AND L. WANTCHEKON (2011): "The slave trade and the origins of mistrust in Africa," *American Economic Review*, 101, 3221–52.

- OSTROM, E. (1990): *Governing the Commons*, Cambridge University Press.
- PAULHUS, D. L. (1984): "Two-component models of socially desirable responding," *Journal of Personality and Social Psychology*, 46, 598.
- PERSSON, T. AND G. TABELLINI (2004): "Constitutions and economic policy," *Journal of Economic Perspectives*, 18, 75–98.
- PUTNAM, R. D., R. LEONARDI, AND R. Y. NANETTI (1993): *Making Democracy Work: Civic Traditions in Modern Italy*, Princeton University Press.
- RAMANKUTTY, N., J. A. FOLEY, J. NORMAN, AND K. MCSWEENEY (2002): "The global distribution of cultivable lands: current patterns and sensitivity to possible climate change," *Global Ecology and Biogeography*, 11, 377–392.
- SHARMA, E., N. MAZAR, A. L. ALTER, AND D. ARIELY (2014): "Financial deprivation selectively shifts moral standards and compromises moral decisions," *Organizational Behavior and Human Decision Processes*, 123, 90–100.
- SHROUT, P. AND N. BOLGER (2002): "Mediation in experimental and nonexperimental studies: New procedures and recommendations," *Psychological Methods*, 7, 422–445.
- SPOLAORE, E. AND R. WACZIARG (2013): "How deep are the roots of economic development?" *Journal of Economic Literature*, 51, 325–69.
- TABELLINI, G. (2008a): "Institutions and culture," *Journal of the European Economic Association*, 6, 255–294.
- (2008b): "The scope of cooperation: Values and incentives," *Quarterly Journal of Economics*, 123, 905–950.
- (2010): "Culture and institutions: economic development in the regions of Europe," *Journal of the European Economic Association*, 8, 677–716.
- USLANER, E. M. (2002): *The Moral Foundations of Trust*, Cambridge University Press.
- USLANER, E. M. AND B. ROTHSTEIN (2016): "The historical roots of corruption: state building, economic inequality, and mass education," *Comparative Politics*, 48, 227–248.
- WEBER, M. (1930): *The Protestant Ethic and the Spirit of Capitalism*, Allen and Unwin.
- WEST, M. D. (2003): "Losers: recovering lost property in Japan and the United States," *Law & Society Review*, 37, 369–424.
- WOOLCOCK, M. (1998): "Social capital and economic development: Toward a theoretical synthesis and policy framework," *Theory and Society*, 27, 151–208.
- ZELLNER, A. (1962): "An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias," *Journal of the American statistical Association*, 57, 348–368.

Appendix A3

Appendices to:
Local Corruption, Income
Underreporting, and Policy
Effectiveness

I. DERIVATION OF THE REGISTRATION MODEL

A. Properties of the Cumulative Binomial Distribution

Let $\mathfrak{B}(m; n, p)$ denote the cumulative binomial distribution that a binary event with probability p occurs at most $m \in \mathbb{N}_0$ out of $n \in \mathbb{N}$ times: $\mathfrak{B}(m; n, p) = \sum_{k=0}^m \binom{n}{k} p^k (1-p)^{n-k}$

Lemma 1 For $m < n$ and $p \in (0, 1)$, $\mathfrak{B}(m; n, p)$ is strictly increasing in m .

Proof. Trivial.

Lemma 2 For $m < n$ and $p \in (0, 1)$, $\mathfrak{B}(m; n, p)$ is strictly decreasing in n .

Proof. Let $B(a, b)$ denote the beta function and $I_x(a, b)$ the regularized incomplete beta function.

$$\begin{aligned} \mathfrak{B}(m; n+1, p) - \mathfrak{B}(m; n, p) &= I_{1-p}(n-m+1, m+1) - I_{1-p}(n-m, m+1) \\ &= I_{1-p}(n-m, m+1) - \frac{(1-p)^n p^{m+1}}{(n-m)B(n-m, m+1)} - I_{1-p}(n-m, m+1) \\ &= -\frac{(1-p)^n p^{m+1}}{\frac{(n-m)!(m!)}{n!}} = -\binom{n}{m} (1-p)^n p^{m+1} < 0 \quad \square \end{aligned} \tag{A3.1}$$

Lemma 3 For $m < n$, $\mathfrak{B}(m; n, p)$ is strictly decreasing in p .

Proof.

$$\begin{aligned} \mathfrak{B}_p(m; n, p) &= \sum_{k=0}^m \binom{n}{k} (kp^{k-1}(1-p)^{n-k} - (n-k)p^k(1-p)^{n-k-1}) \\ &= \frac{n!}{0!(n-0)!} (0 - n(1-p)^{n-1}) \\ &\quad + \frac{n!}{1!(n-1)!} ((1-p)^{n-1} - (n-1)p(1-p)^{n-2}) \\ &\quad + \frac{n!}{2!(n-2)!} (2p(1-p)^{n-2} - (n-2)p^2(1-p)^{n-3}) \\ &\quad + \dots \\ &\quad + \frac{n!}{m!(n-m)!} (mp^{m-1}(1-p)^{n-m} - (n-m)p^m(1-p)^{n-m-1}) \\ &= -\frac{n!}{m!(n-m-1)!} p^m (1-p)^{n-m-1} < 0 \quad \square \end{aligned} \tag{A3.2}$$

B. Deriving the Reporting Function

If a family i with income x_i registers, it will decide what income y_i to report to maximize the expected utility:

$$\max_{y_i \in [0, x_i]} U(y_i | x_i) = \text{Prob}(\text{receiving BF} | y_i) (b - c \cdot (x_i - y_i)) \tag{A3.3}$$

where $b > 0$ is the benefit a family receives from being part of the Bolsa Família program and $c \cdot (x_i - y_i)$ are the expected costs if it is later detected that the family underreported its income.

Claim 1 *If $\bar{x} > \bar{y} + \frac{b}{c} = x^r$, families with an income $x_i \in (x^r, \bar{x}]$ will never report an income $y_i \leq \bar{y}$. Thus, they will never be included in Bolsa Família.*

Proof. The expected utility from reporting y_i given a true income x_i is $Prob(\text{receiving BF}|y_i) (b - c \cdot (x_i - y_i))$. For $x_i > \bar{y} + \frac{b}{c}$, the expected costs of detection exceed the gains from being included in Bolsa Família. As $Prob(\text{receiving BF}|y_i) > 0$ for $y_i \leq \bar{y}$, a family with $x_i > \bar{y} + \frac{b}{c}$ has strictly negative expected utility from reporting an eligible income. \square

For convenience, these families are assumed to report their income truthfully. However, none of the results are affected by this assumption.

Claim 2 *Suppose that families are not constraint to reporting positive incomes. Then, families with an income of $x_i \leq x^r = \bar{y} + \frac{b}{c}$ report according to the unconstrained reporting function $\hat{y}(x_i)$:*

$$\hat{y}(x_i) = x_i - \frac{b}{c} + \frac{1}{\mathfrak{B}(M-1; N-1, F(x_i))} \int_{x_i}^{x^r} \mathfrak{B}(M-1; N-1, F(\alpha)) d\alpha \quad (\text{A3.4})$$

where $\mathfrak{B}(m; n, p)$ denotes the cumulative binomial distribution that a binary event with probability p occurs at most m out of n times.

Proof. The expected utility from reporting y_i given a true income x_i is:

$$\begin{aligned} U(y_i|x_i) &= Prob(\text{receiving BF}|y_i) (b - c \cdot (x_i - y_i)) \\ &= \sum_{k=0}^{M-1} \binom{N-1}{k} F(\hat{y}^{-1}(y_i))^k (1 - F(\hat{y}^{-1}(y_i)))^{N-k-1} (b - c \cdot (x_i - y_i)) \\ &= \mathfrak{B}(M-1; N-1, F(\hat{y}^{-1}(y_i))) (b - c \cdot (x_i - y_i)) \end{aligned} \quad (\text{A3.5})$$

The first order condition with respect to y_i is given by:

$$\begin{aligned} 0 &= \frac{d}{dy_i} [\mathfrak{B}(M-1; N-1, F(\hat{y}^{-1}(y_i))) (b - c \cdot (x_i - y_i))] \\ &= \mathfrak{B}_p(M-1; N-1, F(\hat{y}^{-1}(y_i))) f(\hat{y}^{-1}(y_i)) \frac{1}{\hat{y}'(\hat{y}^{-1}(y_i))} (b - c \cdot (x_i - y_i)) \\ &\quad + c \cdot \mathfrak{B}(M-1; N-1, F(\hat{y}^{-1}(y_i))) \end{aligned} \quad (\text{A3.6})$$

Using $y_i = \hat{y}(x_i)$ and rearranging:

$$\begin{aligned} \frac{d}{dx} [\mathfrak{B}(M-1; N-1, F(x_i)) \hat{y}(x_i)] &= \frac{d}{dx} [\mathfrak{B}(M-1; N-1, F(x_i)) x_i] \\ &\quad - \frac{d}{dx} [\mathfrak{B}(M-1; N-1, F(x_i))] \frac{b}{c} - \mathfrak{B}(M-1; N-1, F(x_i)) \end{aligned} \quad (\text{A3.7})$$

By integrating from x_i to x^r and using the boundary condition $\hat{y}(x^r) = \bar{y}$, the optimal reporting function is recovered:

$$\hat{y}(x_i) = x_i - \frac{b}{c} + \frac{1}{\mathfrak{B}(M-1; N-1, F(x_i))} \int_{x_i}^{x^r} \mathfrak{B}(M-1; N-1, F(\alpha)) d\alpha \quad (\text{A3.8})$$

Next it is verified that no family can increase its expected utility by deviating from the reporting function. Consider the expected utility of a family with a true income x_i that reports as if its income were x' , and suppose all other families report according to $\hat{y}(x)$:

$$\begin{aligned} U(\hat{y}(x')|x_i) &= \mathfrak{B}(M-1; N-1, F(x')) (b - c \cdot (x_i - \hat{y}(x'))) \\ &= \mathfrak{B}(M-1; N-1, F(x')) (b - c \cdot (x' - \hat{y}(x'))) \\ &\quad - c \cdot \mathfrak{B}(M-1; N-1, F(x')) (x_i - x') \\ &= c \cdot \int_{x'}^{x^r} \mathfrak{B}(M-1; N-1, F(\alpha)) d\alpha - c \cdot \int_{x'}^{x_i} \mathfrak{B}(M-1; N-1, F(x')) d\alpha \\ &= c \cdot \int_{x_i}^{x^r} \mathfrak{B}(M-1; N-1, F(\alpha)) d\alpha \\ &\quad + c \cdot \int_{x'}^{x_i} (\mathfrak{B}(M-1; N-1, F(\alpha)) - \mathfrak{B}(M-1; N-1, F(x'))) d\alpha \\ &= \mathfrak{B}(M-1; N-1, F(x_i)) (b - c \cdot (x_i - \hat{y}(x_i))) \\ &\quad + c \cdot \int_{x'}^{x_i} (\mathfrak{B}(M-1; N-1, F(\alpha)) - \mathfrak{B}(M-1; N-1, F(x'))) d\alpha \end{aligned} \quad (\text{A3.9})$$

Because the cumulative binomial distribution function is strictly decreasing in the probability (Lemma 3), and $F(x)$ is strictly increasing, $\mathfrak{B}(M-1; N-1, F(x'))$ is decreasing in x . As a result, the final integral is positive for both $x' < x_i$ and $x' > x_i$ and the expected utility is lower than the expected utility from reporting $y(x_i)$. \square

Claim 3 *If $\hat{y}(0) < 0$ and $\frac{b}{c} < \bar{y}$, there exists a critical value $x^* \in (0, \frac{b}{c}]$ such that a family with $x_i = x^*$ is indifferent between reporting $\hat{y}(x^*)$ and reporting $y = 0$.*

Proof. If all families report according to the constraint reporting function $y(x_i)$, the probability of being included in Bolsa Família with a reported income of $y = 0$ depends on the number of other families that also report $y = 0$. If at most $M - 1$ other families report $y = 0$, reporting $y = 0$ guarantees a place in Bolsa Família. If $k \geq M$ other families report $y = 0$, a family is included with probability $\frac{M}{k+1} < 1$ if it reports $y = 0$. For $x^* = 0$, $\text{Prob}(\text{receiving BF}|0) = 1$. For $x^* > 0$, the probability of inclusion if a family reports $y = 0$ is:

$$\begin{aligned} \text{Prob}(\text{receiving BF}|0) &= \mathfrak{B}(M-1; N-1, F(x^*)) \\ &\quad + \sum_{k=M}^{N-1} \binom{N-1}{k} F(x^*)^k (1 - F(x^*))^{N-k-1} \frac{M}{k+1} \end{aligned}$$

$$\begin{aligned}
&= \mathfrak{B}(M-1; N-1, F(x^*)) + \sum_{k=M+1}^N \binom{N-1}{k-1} F(x^*)^{k-1} (1-F(x^*))^{N-k} \frac{M}{k} \\
&= \mathfrak{B}(M-1; N-1, F(x^*)) + \frac{M}{NF(x^*)} \sum_{k=M+1}^N \binom{N}{k} F(x^*)^k (1-F(x^*))^{N-k} \\
&= \mathfrak{B}(M-1; N-1, F(x^*)) + \frac{M}{NF(x^*)} (1 - \mathfrak{B}(M; N, F(x^*)))
\end{aligned} \tag{A3.10}$$

A family with income x^* is indifferent between reporting $\hat{y}(x^*)$ and $y = 0$ if and only if:

$$\begin{aligned}
&\mathfrak{B}(M-1; N-1, F(x^*)) (b - c \cdot (x^* - \hat{y}(x^*))) \\
&= \left(\mathfrak{B}(M-1; N-1, F(x^*)) + \frac{M}{NF(x^*)} (1 - \mathfrak{B}(M; N, F(x^*))) \right) (b - c \cdot x^*)
\end{aligned} \tag{A3.11}$$

For x^* arbitrarily close to 0, the expected utility from reporting $\hat{y}(x^*)$ is strictly smaller than the expected utility from reporting $y = 0$, because $\hat{y}(x^*) < 0$ and the probability of receiving Bolsa Família from reporting $\hat{y}(x^*)$ is smaller than the probability from reporting $y = 0$. At $x^* = \frac{b}{c}$, the expected utility from reporting $\hat{y}(x^*)$ exceeds the expected utility from reporting $y = 0$:

$$\begin{aligned}
&\mathfrak{B}(M-1; N-1, F(x^*)) (c \cdot \hat{y}(x^*)) = c \cdot \int_{\frac{b}{c}}^{x^r} \mathfrak{B}(M-1; N-1, F(\alpha)) d\alpha \\
&\geq \left(\mathfrak{B}(M-1; N-1, F(x^*)) + \frac{M}{NF(x^*)} (1 - \mathfrak{B}(M; N, F(x^*))) \right) \cdot 0 = 0
\end{aligned} \tag{A3.12}$$

By the intermediate value theorem, there exists $x^* \in (0, \frac{b}{c}]$ such that a family with income x^* is indifferent between reporting $\hat{y}(x^*)$ and reporting 0. \square

C. Optimality of the Reporting Function

To show that the constrained reporting function $y(x)$ is a Bayesian Nash Equilibrium, it is shown that if all other families report according to $y(x)$, a family with an income of $x_i < x^*$ has a higher expected utility from reporting 0 than from reporting any $y > 0$ (Claims 4 and 5), and a family with an income of $x^* \leq x_i < x^r$ has a higher utility from reporting according to $\hat{y}(x_i)$ than from reporting any other $y \geq 0$ (Claim 6).

Claim 4 Suppose all other families report according to $y(x)$. Reporting $y = 0$ yields a higher expected utility than reporting $\hat{y}(x^*)$ if and only if $x_i < x^*$.

Proof. The expected utility from reporting $\hat{y}(x^*)$ is linearly decreasing in x_i at a rate of $-\mathfrak{B}(M-1; N-1, F(x^*)) \cdot c$, while the expected utility from reporting $y = 0$ is linearly decreasing in x_i at the steeper rate of $-(\mathfrak{B}(M-1; N-1, F(x^*)) + \frac{M}{NF(x^*)} (1 - \mathfrak{B}(M; N, F(x^*)))) \cdot c$. By definition, the expected utilities are equal at $x_i = x^*$. Thus, for all $x_i < x^*$, reporting $y = 0$ leads to a strictly higher expected utility. \square

Claim 5 Suppose all other families report according to $y(x)$. For $x_i < x^*$, reporting $y = 0$ is better than reporting $y > 0$.

Proof. For $x_i < x^*$, reporting $y > \hat{y}(x^*)$ leads to lower expected utility than reporting $\hat{y}(x^*)$:

$$\begin{aligned}
& \mathfrak{B}(M-1; N-1, F(\hat{y}^{-1}(y))) (b - c \cdot (x_i - y)) \\
&= \mathfrak{B}(M-1; N-1, F(\hat{y}^{-1}(y))) (b - c \cdot (x^* - y)) \\
&\quad + \mathfrak{B}(M-1; N-1, F(\hat{y}^{-1}(y))) (c \cdot (x^* - x_i)) \\
&\leq \mathfrak{B}(M-1; N-1, F(x^*)) (b - c \cdot (x^* - y(x^*))) \\
&\quad + \mathfrak{B}(M-1; N-1, F(x^*)) (c \cdot (x^* - x_i)) \\
&= \mathfrak{B}(M-1; N-1, F(x^*)) (b - c \cdot (x_i - y(x^*)))
\end{aligned} \tag{A3.13}$$

The inequality holds because $\hat{y}(x^*)$ is the best response for x^* and the cumulative binomial distribution is decreasing in the probability (Lemma 3). As shown earlier, reporting $y = 0$ gives higher expected utility than reporting $\hat{y}(x^*)$ for $x_i < x^*$, thus reporting $y = 0$ is better than reporting any $y > 0$ for $x_i < x^*$. \square

Claim 6 Suppose all other families report according to $y(x)$. For $x^* < x_i < x^r$, reporting $\hat{y}(x_i)$ is better than reporting any other $y \geq 0$.

Proof. Consider the following cases:

1. If $\hat{y}(0) \geq 0$, the reporting function is unconstrained and the claim follows from the optimality of $\hat{y}(x)$.
2. If $\hat{y}(0) < 0$ and $y = 0$, the claim follows from Claim 4.
3. If $\hat{y}(0) < 0$ and $0 < y < \hat{y}(x^*)$, the probability of being included is the same whether the family reports y or $\hat{y}(x^*)$, but the expected utility is strictly smaller if it reports y :
 $\mathfrak{B}(M-1; N-1, F(x^*)) (b - c \cdot (x_i - y)) < \mathfrak{B}(M-1; N-1, F(x^*)) (b - c \cdot (x_i - y(x^*)))$
4. If $\hat{y}(0) < 0$ and $\hat{y}(x^*) \leq y$, the claim follows directly from the optimality of $\hat{y}(x)$.

This concludes the proof. \square

D. Implications of the Reporting Function

Claim 7 For $\frac{b}{c} < \min(\bar{y}, \bar{x})$, at least some families report a strictly positive income $0 < y \leq \bar{y}$.

Proof. The claim follows immediately from Claim 3 because the critical value $x^* \in [0, \frac{b}{c}]$.

Claim 8 For $\frac{b}{c} > \int_0^{\bar{y} + \frac{b}{c}} \mathfrak{B}(M-1; N-1, F(x^*)) d\alpha$, $y(x)$ is discontinuous at x^* .

Proof. Proof by contradiction: Note that if $\frac{b}{c} > \int_0^{\bar{y} + \frac{b}{c}} \mathfrak{B}(M-1; N-1, F(x^*)) d\alpha$ then $\hat{y}(0) < 0$. If the reporting function was continuous at x^* it would need to be the case that $\hat{y}(x^*) = 0$. From Equation (A3.11) it is immediately clear that this also requires that the probabilities of inclusion are the same in the constrained and the unconstrained model. This is only true for $x^* = 0$, contradicting the initial assumption. \square

Claim 9 *If the benefit b increases, more families will report $y = 0$.*

Proof. Consider a family with income x^* . If b increases, the expected utility from reporting $y = 0$ increases more than the expected utility from reporting $\hat{y}(x^*)$. The family now strictly prefers to report $y = 0$. As only families with an income below the critical value report $y = 0$ (Claim 4), the new solution to Equation (A3.11) must be higher than the initial x^* . \square

Claim 10 *If the expected cost c increases, fewer families will report $y = 0$.*

Proof. Consider a family with income x^* . If c increases, the expected utility from reporting $y = 0$ decreases more than the expected utility from reporting $\hat{y}(x^*)$. The family now strictly prefers to report $\hat{y}(x^*)$. As only families with an income above the critical value report according to $\hat{y}(x)$ (Claim 6), the new solution to Equation (A3.11) must be lower than the initial x^* . \square

Claim 11 *For $\bar{x} > x^r$, more families will report $y \leq \bar{y}$ if the expected benefit b increases.*

Proof. The claim follows immediately from the fact that families report an income below the eligibility threshold whenever $x_i < x^r$. \square

Claim 12 *For $\bar{x} > x^r$, fewer families will report $y \leq \bar{y}$ if the expected cost c increases.*

Proof. The claim follows immediately from the fact that families report an income below the eligibility threshold whenever $x_i < x^r$. \square

II. ROBUSTNESS

The first main result of the paper—that Bolsa Família is more effective after a municipality has been audited at random—is robust for a large number of alternative specifications.

A. Following Children for More Than Two Years

The standard estimation follows each child for just two years to mitigate the problem of included families dropping out of Bolsa Família or unincluded families gaining access to the program. This might overestimate the impact of the random audits if the beneficial effect decreases once families have been included for some time.

Table A3.1 displays the estimates of an intent-to-treat approach where families are followed for an additional year and (potentially non-random) dropout and inclusions are ignored. As expected, this leads to lower estimates of Bolsa Família’s effectiveness. However, the gains after a municipality has been audited are similar, if anything, the interaction effects are slightly larger.

TABLE A3.1
BOLSA FAMÍLIA IS MORE EFFECTIVE AFTER A RANDOM AUDIT
(FOLLOWING CHILDREN FOR UP TO THREE YEARS)

	School enrollment (%) (All marginal families)		School enrollment(%) (Newly registered families)		School enrollment (%) (Same registration month)	
	(1)	(2)	(3)	(4)	(5)	(6)
BF	0.583*** (0.045)	0.442*** (0.084)	0.784*** (0.115)	0.521** (0.174)	0.377*** (0.060)	0.239* (0.101)
Past audit		-0.588* (0.245)		-0.913** (0.340)		-0.284 (0.422)
BF × Past audit		0.366* (0.182)		0.692* (0.332)		0.355+ (0.203)
Control mean	87.920	87.920	87.539	87.539	87.631	87.631
Child FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × strata FE	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.858	0.858	0.861	0.861	0.863	0.863
N(municipalities)	5,401	5,401	5,071	5,071	4,860	4,860
N(priority strata)	12,752	12,752	8,803	8,803	6,055	6,055
N(children)	2,585,404	2,585,404	594,973	594,973	751,123	751,123
N	7,171,463	7,171,463	1,657,313	1,657,313	2,095,969	2,095,969
Years	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017

Notes. This table reports the effect of inclusion into the Bolsa Família program and random audits on children’s school enrollment if children are followed for up to three years. The dependent variable in all models takes on value 100 if a child is enrolled in school and 0 otherwise. “BF” indicates if a child’s family is included in the Bolsa Família program, “Past audit” indicates that a municipality has been audited at random, and “BF × Past audit” is the interaction of the two treatments. Columns (1) and (2) present the results for the most representative sample. Columns (3) and (4) consider only newly registered families. Columns (5) and (6) require families to have last updated their data in the same month. All models include individual-level child fixed effects, municipality fixed effects and “Year × Priority strata” fixed effects. For the fixed effects, families in each priority stratum have the exact same income, the same number of children, belong to the same vulnerability category, and have last updated their data in the same month. Standard errors are clustered at both the family and the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

B. Non-Parametric Controls for Age and Gender

Educational participation varies by age and gender. Table A3.2 shows that the result is robust if, in addition to the individual fixed effects, non-parametric controls for age and gender are included and Equation (3.4) is appended as follows:

$$Y_{i,f,\theta,m,t} = \beta \text{ Bolsa Família}_{f,t} + \gamma \text{ Past audit}_{m,t} + \delta (\text{BF} \times \text{Past audit})_{f,m,t} + \alpha_i + \text{age}_{i,t} \times \text{sex}_i + \nu_m + \mu_{\theta,t} + \varepsilon_{i,f,\theta,m,t} \quad (\text{A3.14})$$

TABLE A3.2
BOLSA FAMÍLIA IS MORE EFFECTIVE AFTER A RANDOM AUDIT
(NON-PARAMETRIC CONTROLS FOR AGE AND GENDER)

	School enrollment (%) (All marginal families)		School enrollment(%) (Newly registered families)		School enrollment (%) (Same registration month)	
	(1)	(2)	(3)	(4)	(5)	(6)
BF	1.090*** (0.052)	0.992*** (0.071)	1.524*** (0.119)	1.353*** (0.143)	1.120*** (0.068)	1.024*** (0.092)
Past audit		-0.114 (0.265)		-0.375 (0.366)		0.375 (0.541)
BF × Past audit		0.258* (0.128)		0.454* (0.229)		0.251+ (0.148)
Control mean	87.283	87.283	86.442	86.442	86.809	86.809
Child FE	Yes	Yes	Yes	Yes	Yes	Yes
Age × Sex FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × strata FE	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.928	0.928	0.934	0.934	0.934	0.934
N(municipalities)	5,401	5,401	5,068	5,068	4,858	4,858
N(priority strata)	12,559	12,559	8,641	8,641	6,008	6,008
N(children)	2,573,117	2,573,117	590,630	590,630	747,786	747,786
N	5,146,234	5,146,234	1,181,260	1,181,260	1,495,572	1,495,572
Years	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017

Notes. This table reports the effect of inclusion into the Bolsa Família program and random audits on children's school enrollment if non-parametric controls for age and gender are included. The dependent variable in all models takes on value 100 if a child is enrolled in school and 0 otherwise. "BF" indicates if a child's family is included in the Bolsa Família program, "Past audit" indicates that a municipality has been audited at random, and "BF × Past audit" is the interaction of the two treatments. Columns (1) and (2) present the results for the most representative sample. Columns (3) and (4) consider only newly registered families. Columns (5) and (6) require families to have last updated their data in the same month. All models include individual-level child fixed effects, "Age × Sex" fixed effects, and "Year × Priority strata" fixed effects. For the fixed effects, families in each priority stratum have the exact same income, the same number of children, belong to the same vulnerability category, and have last updated their data in the same month. Standard errors are clustered at both the family and the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

C. Including Teenagers

Youths above the age of 15 were not originally covered by Bolsa Família. While there have been benefits for those aged 16 and 17 during all the years of the analysis, they are subject to different conditionalities, including lower attendance requirements (75% instead of the usual 85%). Moreover, while regular employment is only legal from age 17, apprenticeship contracts are possible from age 15, after the end of compulsory education.

Table A3.3 shows that although the effect of Bolsa Família is somewhat smaller (Columns 1, 3, and 5), the program is still estimated to be significantly more effective after a municipality has been audited at random in the first two samples (Columns 2 and 4), but not in the third one ($P = 0.103$).

TABLE A3.3
BOLSA FAMÍLIA IS MORE EFFECTIVE AFTER A RANDOM AUDIT
(INCLUDING TEENAGERS)

	School enrollment (%) (All marginal families)		School enrollment(%) (Newly registered families)		School enrollment (%) (Same registration month)	
	(1)	(2)	(3)	(4)	(5)	(6)
BF	0.839*** (0.047)	0.750*** (0.063)	1.325*** (0.109)	1.152*** (0.133)	0.765*** (0.059)	0.682*** (0.079)
Past audit		-0.095 (0.255)		-0.364 (0.341)		0.324 (0.540)
BF × Past audit		0.231* (0.113)		0.458* (0.214)		0.213 (0.131)
Control mean	87.214	87.214	84.884	84.884	86.519	86.519
Child FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × strata FE	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.929	0.929	0.941	0.941	0.936	0.936
N(municipalities)	5,416	5,416	5,146	5,146	4,944	4,944
N(priority strata)	15,250	15,250	9,944	9,944	7,205	7,205
N(children)	3,058,499	3,058,499	682,427	682,427	900,544	900,544
N	6,116,998	6,116,998	1,364,854	1,364,854	1,801,088	1,801,088
Years	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017

Notes. This table reports the effect of inclusion into the Bolsa Família program and random audits on children's school enrollment if teenagers up to the age of 17 are included. The dependent variable in all models takes on value 100 if a child is enrolled in school and 0 otherwise. "BF" indicates if a child's family is included in the Bolsa Família program, "Past audit" indicates that a municipality has been audited at random, and "BF × Past audit" is the interaction of the two treatments. Columns (1) and (2) present the results for the most representative sample. Columns (3) and (4) consider only newly registered families. Columns (5) and (6) require families to have last updated their data in the same month. All models include individual-level child fixed effects and "Year × Priority strata" fixed effects. For the fixed effects, families in each priority stratum have the exact same income, the same number of children, belong to the same vulnerability category, and have last updated their data in the same month. Standard errors are clustered at both the family and the municipality level. Significance levels: $^+P < 0.1$, $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

D. Only Original Random Audit Program

In 2016, the CGU was formally reconstituted as the Ministério da Transparência, Fiscalização e Controladoria-Geral da União and the original random audit program, the Programa de Fiscalização por Sorteios Públicos, was superseded by the Programa de Fiscalização em Entes Federativos, that includes both random and non-random audits.

Table A3.4 shows that the results are robust to if only the 40 rounds of the Programa de Fiscalização por Sorteios Públicos are considered, and the random third cycle of the Programa de Fiscalização em Entes Federativos is excluded. This is unsurprising given that only one round of random audits is excluded, affecting the classification of only 35 municipalities in 2017 in the most representative sample, and only 34 and 32 municipalities, respectively, in the smaller samples.

TABLE A3.4
BOLSA FAMÍLIA IS MORE EFFECTIVE AFTER A RANDOM AUDIT
(ONLY ORIGINAL RANDOM AUDIT PROGRAM)

	School enrollment (%) (All marginal families)		School enrollment(%) (Newly registered families)		School enrollment (%) (Same registration month)	
	(1)	(2)	(3)	(4)	(5)	(6)
BF	1.006*** (0.053)	0.897*** (0.072)	1.498*** (0.122)	1.304*** (0.148)	0.919*** (0.067)	0.811*** (0.091)
Past audit		-0.043 (0.352)		-0.343 (0.544)		0.468 (0.698)
BF × Past audit		0.287* (0.131)		0.517* (0.239)		0.279+ (0.152)
Control mean	87.283	87.283	86.441	86.441	86.810	86.810
Child FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × strata FE	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.925	0.925	0.932	0.932	0.931	0.931
N(municipalities)	5,401	5,401	5,068	5,068	4,858	4,858
N(priority strata)	12,559	12,559	8,641	8,641	6,008	6,008
N(children)	2,573,117	2,573,117	590,630	590,630	747,786	747,786
N	5,146,234	5,146,234	1,181,260	1,181,260	1,495,572	1,495,572
Years	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017

Notes. This table reports the effect of inclusion into the Bolsa Família program and random audits on children's school enrollment if only audits under the original Programa de Fiscalização por Sorteios Públicos are considered. The dependent variable in all models takes on value 100 if a child is enrolled in school and 0 otherwise. "BF" indicates if a child's family is included in the Bolsa Família program, "Past audit" indicates that a municipality has been audited at random, and "BF × Past audit" is the interaction of the two treatments. Columns (1) and (2) present the results for the most representative sample. Columns (3) and (4) consider only newly registered families. Columns (5) and (6) require families to have last updated their data in the same month. All models include individual-level child fixed effects, municipality fixed effects and "Year × Priority strata" fixed effects. For the fixed effects, families in each priority stratum have the exact same income, the same number of children, belong to the same vulnerability category, and have last updated their data in the same month. Standard errors are clustered at both the family and the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

E. Including Audits in the Same Year

Because the results of the random audits are often only published at the end of the calendar year or even at the beginning of the next year, the *Past audit* indicator has been defined to take value 1 if a municipality has been audited at random in a *previous* year. Moreover, if the audits increase the effectiveness of Bolsa Família because they make it harder for families to misrepresent their income, including audits that occurred later in the year of registration might underestimate their effectiveness.

If municipalities are considered as having been audited in the past even if the audit takes place in the current year, 198 municipalities are reclassified as having been audited at random earlier than in the original sample. For the smaller samples, 182 and 172 municipalities are affected, respectively. However, Table A3.5 shows that the results don't change. This is also true if only audits from the original Programa de Fiscalização por Sorteios Públicos are considered.

TABLE A3.5
BOLSA FAMÍLIA IS MORE EFFECTIVE AFTER A RANDOM AUDIT
(INCLUDING AUDITS IN THE SAME YEAR)

	School enrollment (%) (All marginal families)		School enrollment(%) (Newly registered families)		School enrollment (%) (Same registration month)	
	(1)	(2)	(3)	(4)	(5)	(6)
BF	1.006*** (0.053)	0.900*** (0.073)	1.498*** (0.122)	1.318*** (0.150)	0.919*** (0.067)	0.807*** (0.092)
Past audit		-0.648** (0.240)		-1.088** (0.347)		-0.469 (0.298)
BF × Past audit		0.275* (0.130)		0.480* (0.237)		0.285+ (0.151)
Control mean	87.277	87.277	86.426	86.426	86.795	86.795
Child FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × strata FE	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.925	0.925	0.932	0.932	0.931	0.931
N(municipalities)	5,401	5,401	5,068	5,068	4,858	4,858
N(priority strata)	12,559	12,559	8,641	8,641	6,008	6,008
N(children)	2,573,117	2,573,117	590,630	590,630	747,786	747,786
N	5,146,234	5,146,234	1,181,260	1,181,260	1,495,572	1,495,572
Years	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017

Notes. This table reports the effect of inclusion into the Bolsa Família program and random audits on children's school enrollment if audits that happen in the same year are considered. The dependent variable in all models takes on value 100 if a child is enrolled in school and 0 otherwise. "BF" indicates if a child's family is included in the Bolsa Família program, "Past audit" indicates that a municipality has been audited at random in a previous year or is being audited at random in the same year, and "BF × Past audit" is the interaction of the two treatments. Columns (1) and (2) present the results for the most representative sample. Columns (3) and (4) consider only newly registered families. Columns (5) and (6) require families to have last updated their data in the same month. All models include individual-level child fixed effects, municipality fixed effects and "Year × Priority strata" fixed effects. For the fixed effects, families in each priority stratum have the exact same income, the same number of children, belong to the same vulnerability category, and have last updated their data in the same month. Standard errors are clustered at both the family and the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

F. Updating the Cadastro Único

Once a family is included in the Cadastro Único, it is required to update its data at least every other year. Thus, in a given year the data of some families actually reflects information from previous years. The identification strategy addressed this challenge by constructing priority strata so that these families would only be matched with other families who also have potentially outdated information. Table A3.6 shows that the results are robust if families in these priority strata are excluded.

TABLE A3.6
BOLSA FAMÍLIA IS MORE EFFECTIVE AFTER A RANDOM AUDIT
(EXCLUDING FAMILIES WITH OLD DATA)

	School enrollment (%) (All marginal families)		School enrollment(%) (Newly registered families)		School enrollment (%) (Same registration month)	
	(1)	(2)	(3)	(4)	(5)	(6)
BF	0.605*** (0.045)	0.509*** (0.062)	0.908*** (0.118)	0.738*** (0.135)	0.588*** (0.060)	0.493*** (0.077)
Past audit		-0.119 (0.293)		-0.229 (0.378)		0.401 (0.600)
BF × Past audit		0.252* (0.119)		0.456* (0.208)		0.247+ (0.138)
Control mean	88.747	88.747	88.188	88.188	87.627	87.627
Child FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × strata FE	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.925	0.925	0.939	0.939	0.933	0.933
N(municipalities)	5,395	5,395	5,045	5,045	4,829	4,829
N(priority strata)	7,233	7,233	5,649	5,649	4,244	4,244
N(children)	2,266,681	2,266,681	496,929	496,929	688,148	688,148
N	4,533,362	4,533,362	993,858	993,858	1,376,296	1,376,296
Years	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017

Notes. This table reports the effect of inclusion into the Bolsa Família program and random audits on children's school enrollment if families that didn't update their data in the year of the matching are excluded. The dependent variable in all models takes on value 100 if a child is enrolled in school and 0 otherwise. "BF" indicates if a child's family is included in the Bolsa Família program, "Past audit" indicates that a municipality has been audited at random, and "BF × Past audit" is the interaction of the two treatments. Columns (1) and (2) present the results for the most representative sample. Columns (3) and (4) consider only newly registered families. Columns (5) and (6) require families to have last updated their data in the same month. All models include individual-level child fixed effects and "Year × Priority strata" fixed effects. For the fixed effects, families in each priority stratum have the exact same income, the same number of children, belong to the same vulnerability category, and have last updated their data in the same month. Standard errors are clustered at both the family and the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

G. Including Only Complete Years

The data for this paper was obtained in late 2017, so the last year of the Cadastro Único data represents the state in June 2017. As a result, families who registered towards the end of the sample had less time to realize their gains, although the year fixed effects mitigate this problem to some degree. Table A3.7 shows that the results are robust if data from 2017 is excluded and only data from complete years are used to estimate the treatment effects.

TABLE A3.7
BOLSA FAMÍLIA IS MORE EFFECTIVE AFTER A RANDOM AUDIT
(EXCLUDING 2017)

	School enrollment (%) (All marginal families)		School enrollment(%) (Newly registered families)		School enrollment (%) (Same registration month)	
	(1)	(2)	(3)	(4)	(5)	(6)
BF	1.026*** (0.054)	0.915*** (0.076)	1.556*** (0.127)	1.349*** (0.160)	0.954*** (0.069)	0.847*** (0.095)
Past audit		-0.044 (0.352)		-0.347 (0.546)		0.468 (0.698)
BF × Past audit		0.292* (0.140)		0.551* (0.280)		0.279+ (0.160)
Control mean	87.152	87.152	86.325	86.325	86.740	86.740
Child FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × strata FE	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.923	0.923	0.928	0.928	0.929	0.929
N(municipalities)	5,381	5,381	5,013	5,013	4,821	4,821
N(priority strata)	11,696	11,696	8,171	8,171	5,734	5,734
N(children)	2,384,393	2,384,393	532,812	532,812	706,958	706,958
N	4,768,786	4,768,786	1,065,624	1,065,624	1,413,916	1,413,916
Years	2012-2016	2012-2016	2012-2016	2012-2016	2012-2016	2012-2016

Notes. This table reports the effect of inclusion into the Bolsa Família program and random audits on children's school enrollment excluding data from 2017 where the data is observed in June instead of December. The dependent variable in all models takes on value 100 if a child is enrolled in school and 0 otherwise. "BF" indicates if a child's family is included in the Bolsa Família program, "Past audit" indicates that a municipality has been audited at random, and "BF × Past audit" is the interaction of the two treatments. Columns (1) and (2) present the results for the most representative sample. Columns (3) and (4) consider only newly registered families. Columns (5) and (6) require families to have last updated their data in the same month. All models include individual-level child fixed effects and "Year × Priority strata" fixed effects. For the fixed effects, families in each priority stratum have the exact same income, the same number of children, belong to the same vulnerability category, and have last updated their data in the same month. Standard errors are clustered at both the family and the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

H. Treatment Propensity

The left panel of Figure A3.1 makes clear that although the overlap assumption is satisfied, families in some priority strata have considerably higher probabilities of being included. Two additional robustness tests suggest themselves from the graph: testing whether treatment effects are robust when inverse probability weights are applied to correct for the higher treatment propensity of some families and when the families with the lowest and highest treatment propensity are excluded.

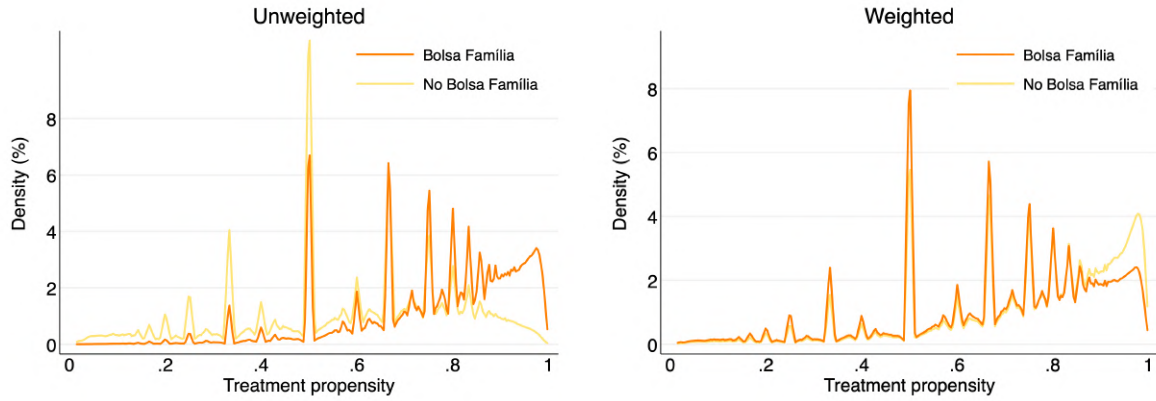


FIGURE A3.1
Overlap of Treatment Propensities

Notes. This figure displays the density functions for the treatment propensities of newly registered families in marginal priority strata. The left panel shows the unweighted density function for Bolsa Família beneficiaries and non-beneficiaries. The right panel shows the density functions after stabilized inverse-probability weights are applied.

The right panel of Figure A3.1 shows that the use of stabilized inverse probability weights does indeed improve the overlap. The weights take the form $w_{1,f} = \frac{Prob(BF)}{Prob(BF|\theta,m)}$ for families that get included and $w_{0,f} = \frac{1-Prob(BF)}{1-Prob(BF|\theta,m)}$ for families that don't get included, where $Prob(BF | \theta, m)$ denotes the conditional probability of being included in Bolsa Família for a family in priority stratum θ and municipality m . Table A3.8 shows that the results are robust: the beneficial effect of Bolsa Família on school enrollment persists if stabilized inverse probability weights are applied (Columns 1, 3, and 5) and, for the most part, continues to be stronger after a municipality has been audited at random (Columns 2 and 4). Only in the sample where families are also required to have registered in the same month is the interaction effect no longer significant ($P = 0.187$).

Table A3.9 shows that the results are relatively robust if only marginal priority strata with a treatment probability of more than 10% and less than 90% are included, although this reduces the sample size considerably. The interaction remains significant in the most representative sample ($P = 0.046$) and the sample of newly registered families ($P = 0.047$), but not in the sample of families who registered in the same month ($P = 0.143$).

TABLE A3.8
BOLSA FAMÍLIA IS MORE EFFECTIVE AFTER A RANDOM AUDIT
(INVERSE PROBABILITY WEIGHTS)

	School enrollment (%) (All marginal families)		School enrollment(%) (Newly registered families)		School enrollment (%) (Same registration month)	
	(1)	(2)	(3)	(4)	(5)	(6)
BF	1.026*** (0.045)	0.922*** (0.067)	1.270*** (0.120)	1.083*** (0.147)	0.928*** (0.070)	0.851*** (0.094)
Past audit		-0.104 (0.292)		-0.277 (0.396)		0.418 (0.661)
BF × Past audit		0.274* (0.128)		0.494* (0.241)		0.200 (0.151)
Control mean	87.283	87.283	86.442	86.442	86.809	86.809
Child FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × strata FE	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.931	0.931	0.937	0.937	0.935	0.935
N(municipalities)	5,401	5,401	5,068	5,068	4,858	4,858
N(priority strata)	12,559	12,559	8,641	8,641	6,008	6,008
N(children)	2,573,117	2,573,117	590,630	590,630	747,786	747,786
N	5,146,234	5,146,234	1,181,260	1,181,260	1,495,572	1,495,572
Years	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017

Notes. This table reports the effect of inclusion into the Bolsa Família program and random audits on children's school enrollment if stabilized inverse probability weights are applied to correct for differences in the treatment propensity across priority strata. The dependent variable in all models takes on value 100 if a child is enrolled in school and 0 otherwise. "BF" indicates if a child's family is included in the Bolsa Família program, "Past audit" indicates that a municipality has been audited at random, and "BF × Past audit" is the interaction of the two treatments. Columns (1) and (2) present the results for the most representative sample. Columns (3) and (4) consider only newly registered families. Columns (5) and (6) require families to have last updated their data in the same month. All models include individual-level child fixed effects, municipality fixed effects and "Year × Priority strata" fixed effects. For the fixed effects, families in each priority stratum have the exact same income, the same number of children, belong to the same vulnerability category, and have last updated their data in the same month. Standard errors are clustered at both the family and the municipality level. Significance levels: $^+P < 0.1$, $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

TABLE A3.9
BOLSA FAMÍLIA IS MORE EFFECTIVE AFTER A RANDOM AUDIT
(FAMILIES WITH TREATMENT PROPENSITY 10–90%)

	School enrollment (%) (All marginal families)		School enrollment(%) (Newly registered families)		School enrollment (%) (Same registration month)	
	(1)	(2)	(3)	(4)	(5)	(6)
BF	0.975*** (0.051)	0.876*** (0.072)	1.483*** (0.112)	1.288*** (0.149)	0.866*** (0.068)	0.781*** (0.092)
Past audit		-0.096 (0.291)		-0.223 (0.509)		0.216 (0.434)
BF × Past audit		0.263* (0.132)		0.526* (0.264)		0.223 (0.152)
Control mean	87.211	87.211	86.494	86.494	86.852	86.852
Child FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × strata FE	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.926	0.926	0.933	0.933	0.932	0.932
N(municipalities)	5,401	5,401	5,060	5,060	4,858	4,858
N(priority strata)	12,547	12,547	8,601	8,601	6,005	6,005
N(children)	2,127,141	2,127,141	470,867	470,867	691,381	691,381
N	4,254,282	4,254,282	941,734	941,734	1,382,762	1,382,762
Years	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017

Notes. This table reports the effect of inclusion into the Bolsa Família program and random audits on children's school enrollment if families in priority strata with less than 10% of more than 90% treatment propensity are excluded. The dependent variable in all models takes on value 100 if a child is enrolled in school and 0 otherwise. "BF" indicates if a child's family is included in the Bolsa Família program, "Past audit" indicates that a municipality has been audited at random, and "BF × Past audit" is the interaction of the two treatments. Columns (1) and (2) present the results for the most representative sample. Columns (3) and (4) consider only newly registered families. Columns (5) and (6) require families to have last updated their data in the same month. All models include individual-level child fixed effects, municipality fixed effects and "Year × Priority strata" fixed effects. For the fixed effects, families in each priority stratum have the exact same income, the same number of children, belong to the same vulnerability category, and have last updated their data in the same month. Standard errors are clustered at both the family and the municipality level. Significance levels: $^+ P < 0.1$, $^* P < 0.05$, $^{**} P < 0.01$, $^{***} P < 0.001$.

III. DETAILS AND RESULTS OF THE FIELD EXPERIMENT

To test if local administrators are indeed less likely to register ineligible families after a random audit, I conducted a field experiment with 6,998 Bolsa Família registration centers (CRAS).¹ Registration centers were contacted asking about the possibility of receiving Bolsa Família, and the information provided in the message was experimentally varied to make the sender eligible or ineligible while holding other characteristics constant.

A. Sample

The sample consisted of 6,998 registration centers that could be contacted individually using email.² There are a total of 8,176 CRAS centers in Brazil, distributed across 5,526 municipalities.³ While the largest municipality, São Paulo, has more than 50 centers, most municipalities (82%) have just one center. The field experiment included only municipalities that were eligible for the random audits. This excluded 557 registration centers located in the 31 most populous municipalities. Of the remaining 7,619 registration centers, 74 were excluded because the official contact list does not include an email address. An additional 547 centers were excluded because they share an email address, making it impossible to contact them individually. This left a final sample of 6,998 registration centers.

B. Treatments

Over several weeks, three emails were sent to registration centers. The emails asked about registering for the Bolsa Família program and provided information that makes the sender either eligible or ineligible (see Table 3.6 of Chapter 3 for the email texts). The "Ineligible" treatment mentioned a monthly income of R\$ 450 and for a mother with one child. The resulting per capita income (R\$ 225) exceeds the eligibility threshold of R\$ 170. The "Eligible I" treatment held the gross monthly income constant but mentioned two children, leading to a per capita income of R\$ 150, which makes the household eligible. Meanwhile, the "Eligible II" treatment held the household composition constant, but reduced the monthly income to R\$ 300, again leading to a per capita income of R\$ 150.

C. Sending the Emails

I set up a private email server to send the emails and collect responses. The use of a custom email server allowed me to ensure that none of the incoming emails are filtered or blocked and

1. The experiment was approved by the Human Subjects Committee of the Faculty of Economics, Business Administration, and Information Technology at the University of Zurich (*OEI IRB # 2019-010*) and was preregistered at the AEA RCT registry under the number *AEARCTR-0004151*.

2. Email addresses are from the official contact details the centers listed on the MDS website: <https://aplicacoes.mds.gov.br/sagi/mops/serv-cras.php> (Accessed on March 21, 2019)

3. Some municipalities do not operate their own center but rely on a neighboring municipality or mobile state-operated services.

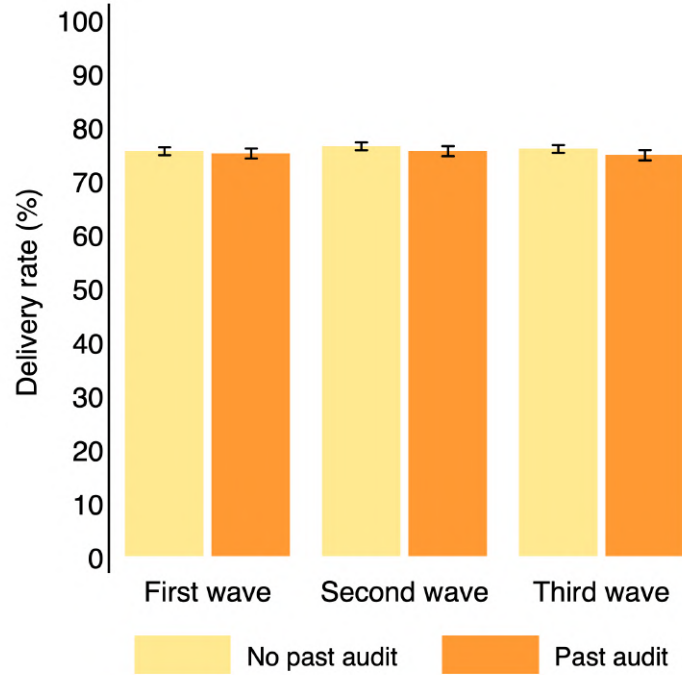


FIGURE A3.2

Email Delivery Rates by Wave and Previous Audit

Notes. This figure displays the delivery rate of the email requests, split by the wave of the experiment and previous audits. Error bars indicate standard errors and are clustered at the municipality level.

that even incorrectly addressed emails would be registered, provided that the domain name was correct. Each outgoing email was sent from a unique email address mentioning only the first name of the sender, *Maria*, and a random five-digit number.

Emails were sent in three waves at the beginning of May, June, and July 2019. Within waves, emails were sent at a random time on a workday between 9:00 and 17:00 in the centers time zone. The order of the emails was randomized at the municipality level and block-randomized with respect to states and whether a municipality had been audited at random. Three different subject lines were used and block-randomly assigned with respect to the treatment, the order of emails, states, and past audit status. Finally, the day of the week and the time of day were block-randomly assigned with respect to all the other design parameters.

Roughly a quarter of emails could not be delivered and returned an error message from the host. Common error messages included not being able to find the user on the host, email memory being full, and timeout errors. This failure rate is independent of the wave of the experiment and whether a municipality had previously been audited or not ($\chi^2(5) = 3.248$, $P = 0.662$; see Figure A3.2), as well as the treatment ($\chi^2(2) = 0.228$, $P = 0.892$).⁴

4. The timing of delivery error messages is also independent of both the treatment and the audit status;

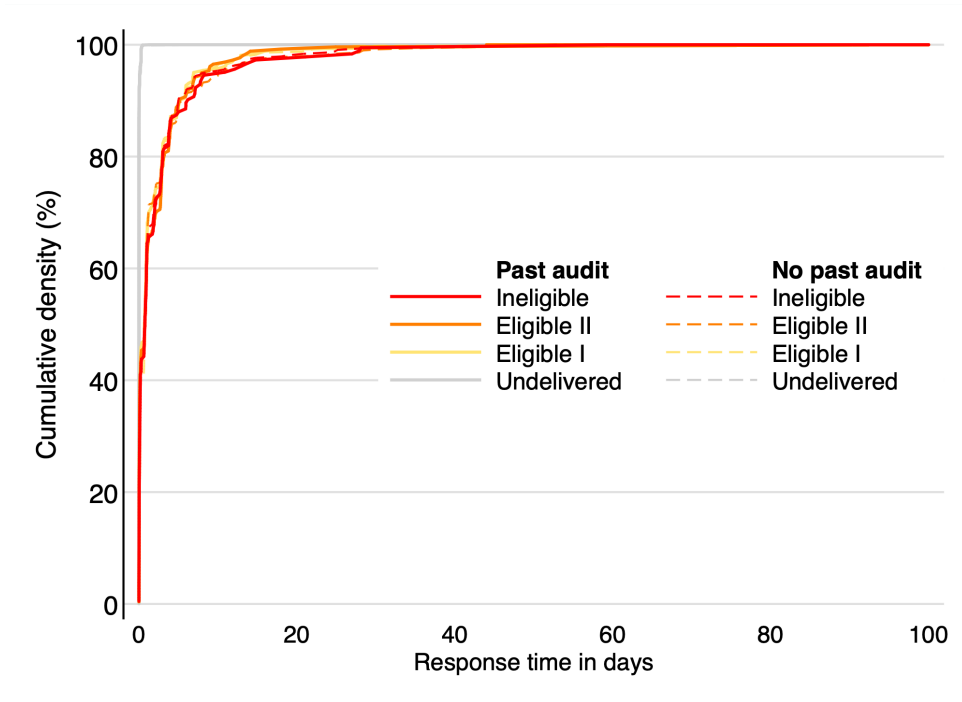


FIGURE A3.3
Response Times by Treatment and Previous Audit

Notes. This figure displays the cumulative distribution function for the time elapsed between the sending of the request and the responses from the registration centers, split by experimental condition and previous audits. The cumulative distribution functions for delivery errors are shown in gray.

D. Response Times

Figure A3.3 displays the cumulative distribution function of the response times. The response times to delivered emails are independent of the treatment and whether a municipality has previously been audited or not; a Kruskal-Wallis rank test does not reject the hypothesis that the response times are drawn from the same distribution ($\chi^2(5) = 1.504$, $P = 0.913$).⁵ I pre-registered that I would use a Heckman selection model to analyze whether the treatments or the interactions of the treatments and the random audits significantly affect the response times. Unsurprisingly, given the Kruskal-Wallis test, they do not.

E. Analysis of Email Content

Even if the response rates in different experimental conditions were the same, employees of registration centers might still discern between eligible and ineligible families when they compose

a Kruskal-Wallis rank test does not reject the hypothesis that the response times are drawn from the same distribution ($\chi^2(5) = 8.148$, $P = 0.148$), suggesting that these are indeed the same automatic server responses.

5. Delivery error messages arrive significantly faster than responses to delivered emails ($\chi^2(1) = 3675.972$, $P = 0.000$).

TABLE A3.10
ANALYSIS OF EMAIL CONTENT
(HECKMAN SELECTION MODEL)

	(1) Incorrect	(2) Correct	(3) Address	(4) Hours	(5) Documents	(6) Call	(7) Rules	(8) Residence
Main								
Ineligible	0.674* (0.299)	0.514 (0.423)	3.408 (2.597)	0.774 (2.145)	1.149 (2.538)	0.069 (1.985)	0.606 (0.751)	2.487 (2.255)
Past audit	-0.003 (0.012)	-0.011 (0.250)	0.833 (2.102)	-2.650 (2.429)	2.135 (2.239)	-2.445 (1.904)	1.162 (0.754)	1.157 (1.944)
Ineligible \times Past audit	-0.741* (0.335)	0.441 (0.874)	-3.977 (4.346)	-1.315 (3.867)	-3.195 (4.038)	7.210+ (4.030)	-1.531 (1.549)	0.187 (3.514)
Response								
Ineligible	-0.082* (0.032)	-0.082* (0.033)	-0.082* (0.033)	-0.082* (0.033)	-0.082* (0.033)	-0.082* (0.033)	-0.082* (0.033)	-0.082* (0.033)
Past audit	0.021 (0.023)	0.021 (0.024)	0.021 (0.024)	0.021 (0.024)	0.021 (0.024)	0.021 (0.024)	0.021 (0.024)	0.021 (0.024)
Ineligible \times Past audit	-0.137** (0.052)	-0.137** (0.052)	-0.137** (0.052)	-0.137** (0.052)	-0.137** (0.052)	-0.137** (0.052)	-0.137** (0.052)	-0.137** (0.052)
Mills λ	0.366 (0.311)	-0.302 (0.908)	-12.215** (4.442)	-8.247* (3.913)	-4.028 (4.231)	-21.450*** (3.526)	-0.733 (1.205)	-13.776*** (2.828)
Control mean	0.000	0.206	43.577	23.330	18.705	13.977	1.542	12.436
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	15891	15891	15891	15891	15891	15891	15891	15891

Notes. This table reports the difference in the content of email responses to requests from eligible and ineligible families in the field experiment. Columns (1) and Columns (2) indicate whether a response contained an incorrect and a correct eligibility assessment, Column (3) whether it provided directions to the CRAS, Column (4) whether it mentions opening hours, Column (5) whether it provides a list of documents needed for the registration, Column (6) whether it asks to call the CRAS, Column (7) whether it mentions the eligibility rules of Bolsa Familia, and Column (8) whether it asks to confirm the place of residence. The dependent variable is an indicator that takes value 100 if the response mentions an item and 0 otherwise. A two-step Heckman selection model is used to account for the differences in the response rate. The upper panel of the table reports the results of the second stage. "Ineligible" indicates that the details in the request made a family ineligible for Bolsa Familia, "Past audit" indicates that a municipality has been audited at random, and "Ineligible \times Past audit" is the interaction of the two treatments. The first stage to predict whether a request receives a response also contained state fixed effects as is reported in the bottom panel of the table. All models include state fixed effects. Bootstrapped standard errors are in parentheses. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

their reply. Thus, in addition to testing for differences in the response rate, each response was also coded to analyze the content of the message.

Table A3.10 test for differences in the content of the emails, using a two-step Heckman selection model to control for the different response rates in the experimental conditions and in previously audited municipalities. As shown in Column (1), responses in the Ineligible treatment were 6.74 percentage points more likely to contain an incorrect assessment of the family's eligibility, i.e., to state that the family qualifies for Bolsa Família ($P = 0.024$). However, this effect is driven exclusively by unaudited municipalities, and it is fully offset by the negative 7.41 percentage point interaction term ($P = 0.027$). Column (2) shows that there is no corresponding effect for correct eligibility assessments. Messages also don't significantly differ in how much practical information they include; they are equally likely to provide directions to the CRAS, to mention opening hours, or to list the documents required for registration. Registration centers in previously audited municipalities are 7.21 percentage points more likely to request that ineligible families call them ($P = 0.074$). While this might open a back channel for colluding with ineligible families, we would expect a negative interaction in this case, as rates of income underreporting are higher in unaudited municipalities. There is also no difference in whether the messages explain Bolsa Família's eligibility criteria or whether they ask the sender to confirm that she lives in the service area of the CRAS. Finally, emails were also screened for direct offers of collusion and for less overt signals of corruption, such as suggesting that the rules are flexible. However, none of the messages contained such a smoking gun.

F. Low Response Rate

The response rate was below 20% for all treatments in all waves of the experiment. There are several possible explanations for this unexpectedly low response rate with different implications: CRAS centers could not have received the emails, they could not be used to correspond via email, or they might have been suspicious.

The relatively high number of emails that returned an error message casts doubt on the reliability of the centers' IT systems and it is conceivable that there are additional emails that did not reach the centers. In the absence of an error message, it is unfortunately not possible to know whether this is the case. However, given that audited and unaudited municipalities were equally likely to return an error message and had similar response rates in the control treatments, it is unlikely that these additional undelivered emails (should they exist) bias the result about the treatment differences.

Not all CRAS centers list an email address as part of their contact details and it is possible that even some that do check their email only infrequently. Especially in rural areas, potential Bolsa Família recipients are arguably more likely to visit the center or to call rather than sending an email.⁶ This factor is very likely to have depressed the response rate. Of the email addresses

6. The use of email was guided in equal parts by financial and practical considerations: telephone numbers

listed, quite a few appear to be private email addresses of employees (some Gmail or Hotmail address) and a small number of responses mentioned that the request had been forwarded to the CRAS by an employee who no longer worked there. Apart from the shift in levels, unfamiliarity with email per se is again unlikely to bias the results, for the same reason outlined above.

The most serious concern is that the low response rate is indicative of suspicion. Several aspects of the design might have contributed to it: the emails stated the sender's approximate income, they provided relatively little other information and no full name, and the email addresses included only the first name. Mentioning the income was unavoidable to make sure that the sender is objectively eligible or ineligible⁷ and providing as little demographic information as possible was an ethical necessity to avoid confusion with real families that register during the time of the study.⁸ It is possible that CRAS employees were more suspicious of emails in the Ineligible treatment and even more so in audited municipalities. If this is the case, it might explain the observed differences in the response rates. Note that it is entirely consistent with the income underreporting mechanism if having previously been audited makes CRAS employees more careful not to register ineligible families because they suspect some audit or test.

G. Robustness

The results of the field experiment are robust for different specifications of the regression, if undelivered emails are coded as non-responses, if only the first wave of emails is considered, and if the two control treatments—Eligible I and Eligible II—are treated separately.

Table A3.11 shows that treating delivery errors as non-responses does not affect the results. As in Table 3.7 of Chapter 3, Column (1) shows the results if state fixed effects are included to account for the stratification of the audit lottery, Columns (2) and (3) show the results if only within CRAS center variance is exploited, and Columns (4) and (5) if the control variables from Avis et al. (2018) are used. Columns (3) and (5) again add fixed effects for the different subject lines, the order, and the timing of the emails. Although the coefficients are somewhat smaller when undelivered emails are included, they remain statistically significant, as is the case for the interaction terms ($P = 0.001$ for the coefficients and $P \leq 0.020$ for the interactions in all specifications).

in Brazil are geographically coded and might have revealed that the caller does not reside in the municipality.

7. Other approaches such as mentioning an occupation would have required CRAS employees to guess whether the household may or may not be eligible and would have depended on local factors such as the median wage.

8. The household composition—a single mother with one or two children—is very common among registrants, there is no information on the children's names, age or gender, and Maria is by far the most common female name in Brazil. Almost 15% of women in the Cadastro Único list Maria as one of their first names. Even if an administrator thought that the email matches someone he knows, it would be very easy to argue that it is just a coincident.

TABLE A3.11
RESPONSE RATE TO REQUESTS FROM ELIGIBLE AND INELIGIBLE
FAMILIES
(INCLUDING UNDELIVERED EMAILS)

	(1)	(2)	(3)	(4)	(5)
Ineligible	-1.284** (0.450)	-1.284** (0.450)	-1.304** (0.452)	-1.284** (0.450)	-1.349** (0.450)
Past audit	0.392 (0.785)			0.168 (0.747)	0.171 (0.745)
Ineligible × Past audit	-1.752* (0.709)	-1.752* (0.709)	-1.670* (0.709)	-1.752* (0.709)	-1.684* (0.714)
Population (Log.)				2.554*** (0.401)	2.576*** (0.401)
Income inequality (Gini)				-8.170+ (4.941)	-8.179+ (4.965)
Income per capita (Log.)				4.993*** (1.425)	4.875*** (1.428)
Illiteracy				0.015 (0.058)	0.018 (0.058)
Urban population				-0.647 (1.794)	-0.574 (1.800)
Control mean	10.864	10.864	10.864	10.864	10.864
State FE	Yes	Yes	Yes	Yes	Yes
Center FE	No	Yes	Yes	No	No
Order FE	No	No	Yes	No	Yes
Subject line FE	No	No	Yes	No	Yes
Day FE	No	No	Yes	No	Yes
Time FE	No	No	Yes	No	Yes
R2	0.037	0.600	0.606	0.058	0.067
N	20994	20994	20994	20994	20994

Notes. This table reports the difference in response rates to requests from eligible and ineligible families in the field experiment if undelivered emails are coded as non-responses. The dependent variable is an indicator that takes value 100 if the registration center replied to a request and 0 otherwise. “Ineligible” indicates that the details in the request made a family ineligible for Bolsa Família, “Past audit” indicates that a municipality has been audited at random, and “Ineligible × Past audit” is the interaction of the two treatments. “Population (Log.)”, “Income inequality (Gini)”, “Income per capita (Log.)”, and the rates of “Illiteracy” and “Urban population” control for municipality characteristics in 2000, before the inception of the audits program. Columns (2) and (3) include registration center fixed effects. Columns (3) and (5) include fixed effects for the order of emails, the different subject lines, the day of the week and the exact time of day the email was sent. All models include state fixed effects. Standard errors are clustered at the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

As each center receives three similar emails, the response to emails in later waves might be affected by the emails in earlier waves: the email might look familiar to social workers or be more likely to end up in a spam filter.⁹ To address these concerns, I pre-registered a robustness check to show that the effect persists if only the first wave of emails is used. Table A3.12 shows that

9. The response messages were screened for signs of suspicion. Overall, fewer than 1% of the responses showed any sign of suspicion. However, this rate increases with each round from 0.27% in the first round to 1.52% in the final round.

TABLE A3.12
RESPONSE RATE TO REQUESTS FROM ELIGIBLE
AND INELIGIBLE FAMILIES
(FIRST WAVE ONLY)

	(1)	(2)	(3)
Ineligible	-2.174 ⁺ (1.245)	-1.904 (1.232)	-1.802 (1.247)
Past audit	1.651 (1.402)	1.536 (1.342)	1.386 (1.371)
Ineligible \times Past audit	-4.422* (1.902)	-4.692* (1.890)	-4.743* (1.947)
Population (Log.)		2.868*** (0.608)	2.954*** (0.612)
Income inequality (Gini)		-11.944 (8.903)	-13.252 (8.987)
Income per capita (Log.)		5.686* (2.319)	5.336* (2.363)
Illiteracy		-0.001 (0.093)	-0.002 (0.096)
Urban population		1.060 (3.047)	1.183 (3.093)
Control mean	15.064	15.064	15.064
State FE	Yes	Yes	Yes
Subject line FE	No	No	Yes
Day FE	No	No	Yes
Time FE	No	No	Yes
R ²	0.043	0.066	0.098
N	5276	5276	5276

Notes. This table reports the difference in response rates to requests from eligible and ineligible families in the first round of the field experiment. The dependent variable is an indicator that takes value 100 if the registration center replied to a request and 0 otherwise. “Ineligible” indicates that the details in the request made a family ineligible for Bolsa Família, “Past audit” indicates that a municipality has been audited at random, and “Ineligible \times Past audit” is the interaction of the two treatments. “Population (Log.)”, “Income inequality (Gini)”, “Income per capita (Log.)”, and the rates of “Illiteracy” and “Urban population” control for municipality characteristics in 2000, before the inception of the audits program. Column (3) includes fixed effects for the order of emails, the different subject lines, the day of the week and the exact time of day the email was sent. All models include state fixed effects. Robust standard errors are reported in brackets. Significance levels: ⁺ $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

the effect persists. Column (1) shows the model if only state fixed effects are included. Column (2) adds the control variables from Avis et al. (2018). Column (3) further add fixed effects for the different subject lines and the timing of the emails. In all three specifications, emails from the Ineligible treatment were approximately two percentage points less likely to receive a response in unaudited municipalities, and 6.5 percentage points less likely to receive a response in audited municipalities. The treatment effects are significantly bigger in municipalities that have been audited at random ($P < 0.020$ in all specifications).

TABLE A3.13
RESPONSE RATE TO REQUESTS FROM ELIGIBLE AND INELIGIBLE
FAMILIES
(SEPARATE CONTROL TREATMENTS)

	(1)	(2)	(3)	(4)	(5)
Eligible II	0.042 (0.645)	-0.060 (0.648)	-0.234 (0.651)	0.051 (0.645)	-0.091 (0.649)
Ineligible	-1.687* (0.672)	-1.556* (0.682)	-1.617* (0.686)	-1.646* (0.671)	-1.732** (0.672)
Past audit	0.513 (1.180)			0.307 (1.111)	0.363 (1.112)
Eligible II \times Past audit	-0.118 (1.126)	-0.029 (1.127)	-0.095 (1.123)	-0.113 (1.124)	-0.064 (1.126)
Ineligible \times Past audit	-2.455* (1.094)	-2.717* (1.105)	-2.740* (1.107)	-2.492* (1.093)	-2.545* (1.103)
Population (Log.)				3.510*** (0.488)	3.509*** (0.489)
Income inequality (Gini)				-11.855+ (6.140)	-12.169* (6.168)
Income per capita (Log.)				5.632** (1.765)	5.618** (1.766)
Illiteracy				0.015 (0.075)	0.020 (0.075)
Urban population				-0.707 (2.235)	-0.533 (2.244)
Eligible II - Ineligible	0.010	0.027	0.043	0.012	0.016
Diff. Interactions	0.031	0.013	0.015	0.028	0.023
Control mean	14.277	14.345	14.345	14.277	14.277
State FE	Yes	Yes	Yes	Yes	Yes
Center FE	No	Yes	Yes	No	No
Order FE	No	No	Yes	No	Yes
Subject line FE	No	No	Yes	No	Yes
Day FE	No	No	Yes	No	Yes
Time FE	No	No	Yes	No	Yes
R2	0.039	0.599	0.606	0.068	0.078
N	15891	15736	15736	15891	15891

Notes. This table reports the difference in response rates to requests from eligible and ineligible families in the field experiment if the two control conditions are treated separately. The dependent variable is an indicator that takes value 100 if the registration center replied to a request and 0 otherwise. “Eligible II” indicates that the request is part of the Eligible II condition, “Ineligible” indicates that the details in the request made a family ineligible for Bolsa Família, “Past audit” indicates that a municipality has been audited at random, and “Eligible II \times Past audit” and “Ineligible \times Past audit” are the interactions of the two conditions with the random audits. “Population (Log.)”, “Income inequality (Gini)”, “Income per capita (Log.)”, and the rates of “Illiteracy” and “Urban population” control for municipality characteristics in 2000, before the inception of the audits program. Columns (2) and (3) include registration center fixed effects. Columns (3) and (5) include fixed effects for the order of emails, the different subject lines, the day of the week and the exact time of day the email was sent. All models include state fixed effects. Standard errors are clustered at the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Finally, while both control treatments imply the same per capita income and qualify the household to receive Bolsa Família, they could nevertheless be perceived differently by administrators at the CRAS. I thus test whether the results persist if the Eligible I and Eligible II treatments are included separately in the regression. Table A3.13 shows that this is indeed the case. While there is no significant difference between either the two control treatments or their interactions with previous audits ($p > 0.700$ for all comparisons), the coefficients for the Ineligible treatment and its interaction are quantitatively similar to the values in Table 3.7 of Chapter 3 where they are compared to both control treatments simultaneously. Both the coefficients and the interaction terms differ significantly from those of the Eligible I treatment ($P < 0.020$ for all comparisons) and the Eligible II treatment ($P < 0.040$ for all comparisons).

IV. DETAILS AND RESULTS OF THE SOCIAL NORMS EXPERIMENT

To test whether changes in local norms can explain the lower rates of income underreporting after a random audit, I elicited relevant social norms in an incentivized online experiment with 675 participants living in 424 municipalities, some of which had been randomly selected for audits in the past.¹⁰ However, there is no evidence that social norms change as a result of a random audit.

A. Recruitment and Sample

Participants were recruited through Facebook to achieve maximum geographic coverage and, at the same time, relatively precise targeting. The target audience was restricted to resemble the typical Bolsa Família beneficiary—women aged 18 to 50, who have an interest in Bolsa Família and are accessing Facebook on their mobile device. The 31 municipalities that are too populous for the random audits were excluded. The recruitment advert (Figure A3.4) only mentioned the possibility of winning mobile phone credit and made no mention of Brazil’s audit program.



FIGURE A3.4
Facebook Recruitment Advert (English Translation)

Notes. This figure displays the English translation of the advert used to recruit participants on Facebook.

While Facebook allows advertisers to target relatively homogenous audiences, it does not guarantee that participation is evenly distributed across municipalities. Nor is it guaranteed

10. The experiment was approved by the Human Subjects Committee of the Faculty of Economics, Business Administration, and Information Technology at the University of Zurich (*OEC IRB # 2019-008*).

TABLE A3.14
BALANCEDNESS OF INDIVIDUAL CHARACTERISTICS IN ONLINE SAMPLE

	No past audit		Past audit		LHS-test
	(1)		(2)		(3)
	Mean	Standard Deviation	Mean	Standard Deviation	Coeff.
Population (Log.)	10.464	1.239	10.531	1.231	-0.011
Urban population	0.770	0.212	0.745	0.197	-0.001
Income inequality (Gini)	0.545	0.055	0.561	0.050	0.009 ⁺
Income per capita (Log.)	5.998	0.512	5.874	0.513	0.015
Illiteracy	16.801	11.191	20.017	12.089	0.124
F-test: $\chi^2(5)$					3.344
F-test: P-value					0.647
Observations	271		153		424
Age	30.727	7.891	30.796	7.796	0.176
Education (7-point scale)	3.064	1.355	2.941	1.398	-0.195
Household size	3.778	1.683	3.452	1.441	-0.328*
In Cadastro Único	0.800	0.401	0.860	0.348	0.058
Ever a beneficiary	0.833	0.374	0.887	0.317	0.055
F-test: $\chi^2(5)$					11.604
F-test: P-value					0.041
Observations	454		221		675

Notes. This table reports on the balancedness of municipal (top panel) and individual (bottom panel) characteristics in the online sample. Columns (1) and (2) present the summary statistics for municipalities that have not previously been subject to a random audit and for those that have been, respectively. Column (3) uses a left-hand-side test (Pei et al., 2019) to check whether these characteristics are predictive of whether a municipality has been audited in regressions of the form $X_{i,m,s} = \alpha + \beta \text{Past Audit}_{m,s} + \nu_s + \varepsilon_{i,m,s}$. The F-test tests whether these coefficients are jointly different from zero. Significance levels: ⁺ $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

that participants from municipalities that have previously been audited at random are similar to those from unaudited municipalities along dimensions that cannot directly be targeted.

The top panel of Table A3.14 shows that audited and unaudited municipalities in the online sample do not differ significantly with regard to the socioeconomic factors used elsewhere in this paper. Columns (1) and (2) show that these characteristics are relatively well-balanced between municipalities that have not previously been subject to a random audit and those that have been. As a more stringent test of balancedness (Pei et al., 2019), Column (3) reports the coefficient when the variable of interest is regressed on the audit indicator and state fixed effects, and tests whether these coefficients are jointly significant from zero. A significant test statistic suggests that these variables together can predict which municipalities in the online sample have been audited. This is not the case ($\chi^2(5) = 3.344$, $P = 0.647$).

Turning to individual characteristics, the bottom panel shows that there are some differences between respondents from audited and unaudited municipalities. The households of participants from audited municipalities are significantly smaller (-0.328 , $P = 0.032$), and the LHS-test cannot reject the hypothesis that there are some differences in participant characteristics ($\chi^2(5) = 11.604$, $P = 0.041$). As a robustness test, these variables are included in the following tables.

B. Experimental Design

While society at large condemns welfare fraud, some parts of the population may nonetheless consider it an acceptable practice. However, as welfare fraud is illegal, few participants would admit to supporting it if asked directly. This non-negligible social desirability problem is mitigated using the method developed by Krupka and Weber (2013): Instead of asking participants for their opinion, they are incentivized to try and give the same response as another randomly matched participant. This transforms the question into a coordination game in which the social norm serves as a focal point.

I presented participants with three short vignettes (see Table 3.8 in Chapter 3): First, a family that underreports its income to qualify for the program. Second, a local administrator who suspects that the family underreports their income but turns a blind eye. Finally, a neighbor who places an anonymous call with the local registration office that leads to an audit of the family. Participants are then asked to rate the behavior in the scenario, given four choices: very wrong, somewhat wrong, somewhat right, very right. In each case, participants are asked to try to give the same answer as another participant: "Please indicate how right or wrong [...]’s action is. Both you and the other participant should try to mark the same response."¹¹

Participants were informed that one of the three vignettes will be randomly selected for payment at the end of the experiment and that they receive R\$ 10 in mobile phone credits if the answers match. Given the strong focal point for each of the vignettes, a participant who always gave the most common response had an expected payout of R\$ 7.00. The average participant received just over R\$ 6.06. On top of this, participants had the chance to win an additional R\$ 10 in a Fischbacher and Föllmi-Heusi (2013) honesty task (see below). The median participant took 7 minutes and 15 seconds to complete the experiment, leading to a fairly generous reimbursement on a per hour basis, despite the absence of a participation fee.

In addition to the vignettes, participants were asked to indicate how prevalent behavior similar to the scenarios is in their municipality. Finally, participants also provided basic demographic information.

C. Social Norms

Table A3.15 displays the results of an ordered logistic regression to test whether participants from municipalities that were randomly audited in the past are more likely to reply that the behavior described in the vignettes is wrong. Columns (1), (2), and (3) report the norm on income underreporting, Columns (4), (5), and (6) on helping families to underreport their

11. Anecdotal evidence from participants’ interactions with the Facebook advert suggests that participants understand the incentives and find them hilarious:

Only gets if you say the same answer as the other person 😂😂😂😂😂😂

TABLE A3.15
SOCIAL NORMS DON'T CHANGE AFTER A RANDOM AUDIT
(ORDERED LOGISTIC REGRESSION)

	Income underreporting			Turning a blind eye			Blowing the whistle		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Past audit	-0.078 (0.235)	-0.124 (0.240)	-0.125 (0.246)	0.074 (0.247)	0.083 (0.249)	0.055 (0.255)	-0.113 (0.213)	-0.132 (0.216)	-0.067 (0.218)
Age		0.030* (0.014)	0.030* (0.014)		0.043** (0.015)	0.044** (0.015)		-0.005 (0.011)	-0.006 (0.012)
Education (7-point scale)		0.032 (0.082)	0.040 (0.084)		0.166 ⁺ (0.087)	0.173 ⁺ (0.090)		0.027 (0.078)	0.016 (0.080)
Household size		-0.080 (0.066)	-0.082 (0.067)		0.028 (0.063)	0.025 (0.064)		-0.072 (0.061)	-0.066 (0.063)
In Cadastro Único		0.214 (0.396)	0.224 (0.397)		-0.117 (0.442)	-0.107 (0.448)		0.133 (0.300)	0.108 (0.301)
Ever a beneficiary		-0.111 (0.436)	-0.077 (0.441)		0.059 (0.452)	0.088 (0.455)		-0.051 (0.351)	-0.092 (0.357)
Population (Log.)			-0.065 (0.115)			-0.096 (0.117)			0.129 (0.115)
Income inequality (Gini)			-2.361 (2.977)			1.165 (3.523)			3.147 (2.275)
Income per capita (Log.)			0.710 (0.743)			0.137 (0.830)			-1.237 ⁺ (0.683)
Illiteracy			0.009 (0.030)			-0.012 (0.038)			0.003 (0.026)
Urban population			0.067 (0.845)			0.190 (0.913)			0.678 (0.866)
Control mean	3.544	3.544	3.544	3.540	3.540	3.540	1.526	1.526	1.526
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N(municipalities)	424	424	424	424	424	424	424	424	424
N	675	675	675	675	675	675	675	675	675

Notes. This table reports the difference in elicited social norms in the online experiment. Columns (1), (2), and (3) report the norm on income underreporting, Columns (4), (5), and (6) on helping families to underreport their income, and Columns (7), (8), and (9) on blowing the whistle on families who underreport. The dependent variable is an ordinal scale whether the behavior described in the vignette is: very wrong, somewhat wrong, somewhat right, very right. “Past audit” indicates that a municipality has been audited at random. “Age”, “Education (7-point scale)”, “Household size”, “In Cadastro Único”, and “Ever a beneficiary” control for individual-level characteristics. “Population (Log.)”, “Income inequality (Gini)”, “Income per capita (Log.)”, and the rates of “Illiteracy” and “Urban population” control for municipality characteristics in 2000, before the inception of the audits program. All models include state fixed effects. Robust standard errors are clustered at the municipality level. Significance levels: ⁺ $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

income, and Columns (7), (8), and (9) on denouncing families who underreport. For each vignette, the first specification includes only the audit indicator and state fixed effects to account for the stratification of the audit lottery. The second specification includes participant characteristics and the third both participant and municipality level controls. The coefficient of the audit indicator does not reach statistical significance in any of the models ($P > 0.500$ for all specifications).¹²

12. The result is not specific to the ordered logistic model. The same is true in a linear model where the responses are coded from (1) *very right* to (4) *very wrong*.

TABLE A3.16
BELIEFS DON'T CHANGE AFTER A RANDOM AUDIT
(ORDERED LOGISTIC REGRESSION)

	Income underreporting			Turning a blind eye			Blowing the whistle		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Past audit	-0.034 (0.184)	-0.041 (0.183)	-0.106 (0.187)	-0.197 (0.181)	-0.201 (0.182)	-0.184 (0.176)	-0.086 (0.187)	-0.123 (0.192)	-0.121 (0.189)
Age		0.007 (0.009)	0.005 (0.009)		-0.008 (0.009)	-0.008 (0.009)		-0.005 (0.009)	-0.003 (0.009)
Education (7-point scale)		0.085 (0.054)	0.068 (0.056)		0.026 (0.056)	0.017 (0.058)		-0.099 (0.061)	-0.093 (0.062)
Household size		-0.039 (0.052)	-0.043 (0.052)		0.031 (0.049)	0.042 (0.048)		-0.063 (0.053)	-0.059 (0.053)
In Cadastro Único		0.155 (0.246)	0.106 (0.254)		-0.114 (0.248)	-0.140 (0.262)		-0.251 (0.273)	-0.240 (0.272)
Ever a beneficiary		0.078 (0.271)	0.071 (0.274)		0.520 ⁺ (0.278)	0.478 ⁺ (0.283)		0.222 (0.320)	0.191 (0.321)
Population (Log.)			0.223* (0.087)			0.182 ⁺ (0.098)			-0.037 (0.096)
Income inequality (Gini)			-0.549 (2.010)			-3.630 ⁺ (2.177)			-2.172 (2.282)
Income per capita (Log.)			0.271 (0.524)			-0.447 (0.582)			0.073 (0.570)
Illiteracy			-0.010 (0.026)			-0.008 (0.025)			-0.005 (0.025)
Urban population			-1.960* (0.776)			-0.856 (0.760)			-0.611 (0.723)
Control mean	2.976	2.976	2.976	2.159	2.159	2.159	2.075	2.075	2.075
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N(municipalities)	424	424	424	424	424	424	424	424	424
N	675	675	675	675	675	675	675	675	675

Notes. This table reports the difference in beliefs in audited and unaudited municipalities. Columns (1), (2), and (3) report the beliefs about the frequency of income underreporting, Columns (4), (5), and (6) of helping families to underreport their income, and Columns (7), (8), and (9) of blowing the whistle on families who underreport. The dependent variable is an ordinal scale whether the behavior described in the vignette is: very rarely, rarely, often, very often. “Past audit” indicates that a municipality has been audited at random. “Age”, “Education (7-point scale)”, “Household size”, “In Cadastro Único”, and “Ever a beneficiary” control for individual-level characteristics. “Population (Log.)”, “Income inequality (Gini)”, “Income per capita (Log.)”, and the rates of “Illiteracy” and “Urban population” control for municipality characteristics in 2000, before the inception of the audits program. All models include state fixed effects. Robust standard errors are clustered at the municipality level. Significance levels: ⁺ $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

D. Beliefs

Even if the random audits do not change participants’ moral assessment of income underreporting, they might change the perceived frequency of these behaviors. To address this, the online experiment asked participants how commonly actions similar to that in the scenarios occur in their municipality. Table A3.16 displays the results of an ordered logistic regression, this time with the perceived frequency on an ordinal scale: very rarely, rarely, often, very often. Again, the coefficient of the audit indicator is not statistically significant in any of the models

($P > 0.250$ for all specifications).¹³

E. Rule-Following

To test whether the random audits affect general rule-following, I used a version of the honesty game developed by Fischbacher and Föllmi-Heusi (2013). In the task, a participant flips a coin in private and reports the outcome to the experimenter. Without any verification, the participant is then paid when she reports *heads* as the outcome. Since it is impossible to know whether any particular participant reports the true outcome—or, indeed, whether she even flipped a coin—, each participant has an incentive to maximize the payoff by just reporting *heads*. On aggregate, however, it is easy to determine the share of participants that misreport the outcome of the coin flip because roughly half of the reported coins should come up *tails*.¹⁴

Participants flipped a R\$1 coin ten times and reported the outcome of each round. One coin flip was selected at random at the end of the experiment, and participants won an additional R\$ 10 if they had reported that it came up *heads*.¹⁵ Thus, if the random audits change norms about rule-following behavior in general, we would expect participants in previously audited municipalities to report fewer instances of *heads*.

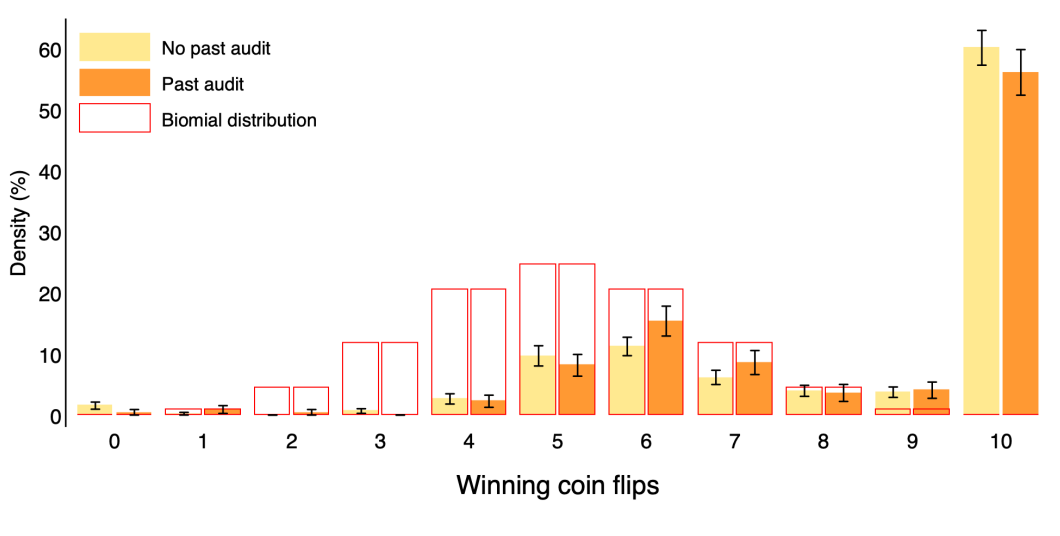


FIGURE A3.5

Distribution of Reported Winning Coin Flips in the Honesty Game

Notes. This figure displays the reported number of winning coin flips in the Fischbacher and Föllmi-Heusi (2013) honesty game. The red outlines indicate the expected distribution of winning coin flips if participants play honestly. Error bars indicate standard errors and are clustered at the municipality level.

13. The same is true in a linear model where the responses are coded from (1) *very rarely* to (4) *very often*.

14. Versions of this task have been successfully used to study honesty in the banking sector (Cohn et al., 2014), to investigate the development of honesty in children (Buccioli and Piovesan, 2011), to demonstrate the effect of experienced unfairness on behavior (Houser et al., 2012), and to predict rule violations in a maximum-security prison (Cohn et al., 2015).

15. This payment scheme was necessary because some mobile phone operators required a minimum of R\$ 10 for credit top-up.

Figure A3.5 shows that this is not the case. While participants from audited municipalities are slightly less likely to report zero (-1.35 percentage points, $P = 0.090$) and somewhat more likely to report six successful coin flips ($+5.67$ percentage points, $P = 0.048$), none of the differences is significant if we correct for multiple hypothesis testing (Benjamini and Hochberg, 1995). In particular, residents in audited municipalities are not significantly less likely to report all coin flips as successful, even without correcting for multiple hypothesis testing ($P = 0.184$). Overall, we cannot reject the hypothesis that the distribution of supposedly winning coin flips is the same for participants from audited and unaudited municipalities ($\chi^2(10) = 10.8348$, $P = 0.371$, χ^2 -test; $Z = 0.675$, $P = 0.500$, rank-sum test).

V. ADDITIONAL FIGURES

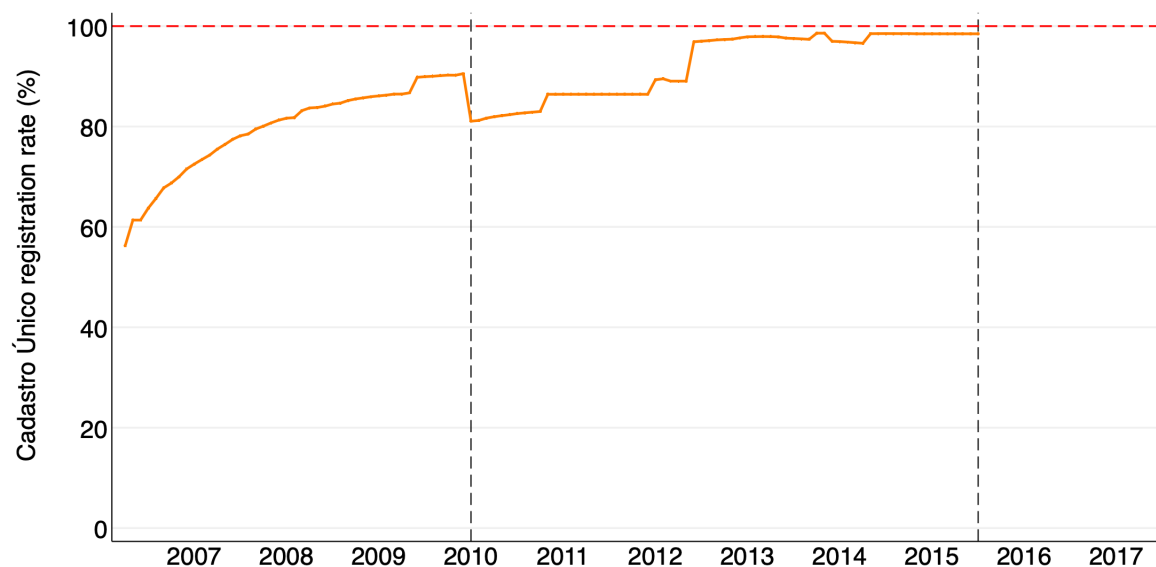


FIGURE A3.6

Cadastro Único Registration Rate

Notes. This figure displays the average Cadastro Único registration rate across time. The index of municipal management quality (IGD-M) defines the registration rate as the number of registered families divided by the number of eligible families in the municipality, estimated based on the last census. The indicator was dropped from the IGD-M in July 2015.

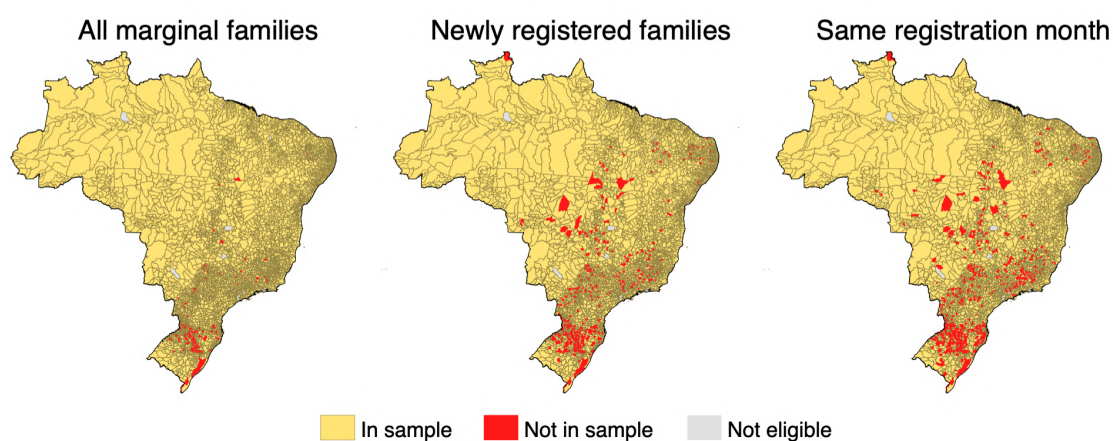


FIGURE A3.7

Geographic Coverage of Marginal Priority Strata

Notes. This figure displays the geographic coverage of marginal priority strata in the most representative sample (left), the sample of newly registered families (middle), and the sample of families that registered in the same month (right). Gray municipalities are not included because they are not eligible for the random audits.

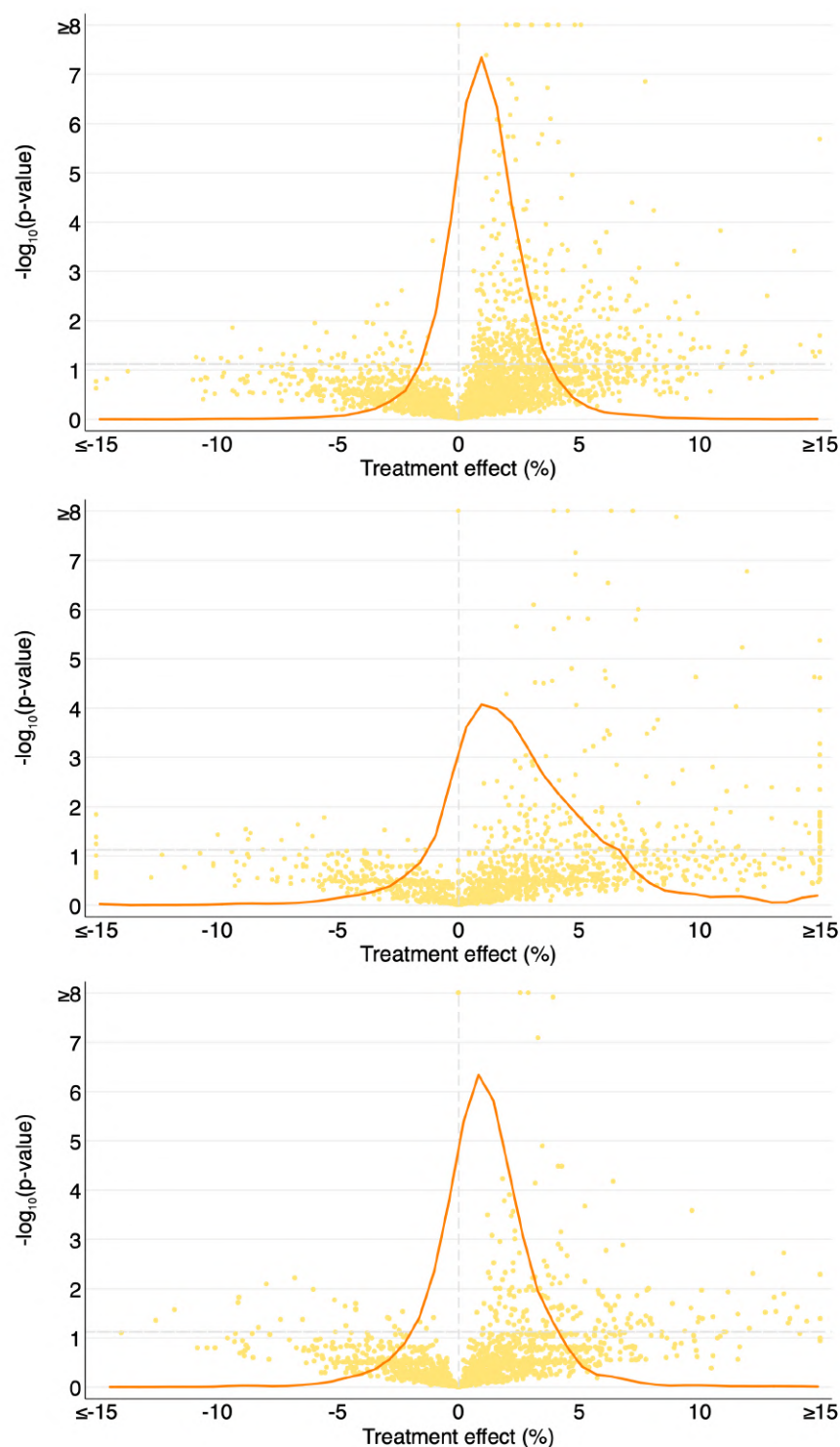


FIGURE A3.8

Distribution of Bolsa Família's Effectiveness in Different Municipalities

Notes. This figure displays the distribution of estimated treatment effects for different municipalities in the most representative sample (top), the sample of newly registered families (middle), and the sample of families that registered in the same month (bottom). Each dot represents the effect size and p-value for a municipality and is estimated based on at least 50 children. The density is weighted by the number of observations in each group. The dashed vertical gray line represents significance at the 5% level.

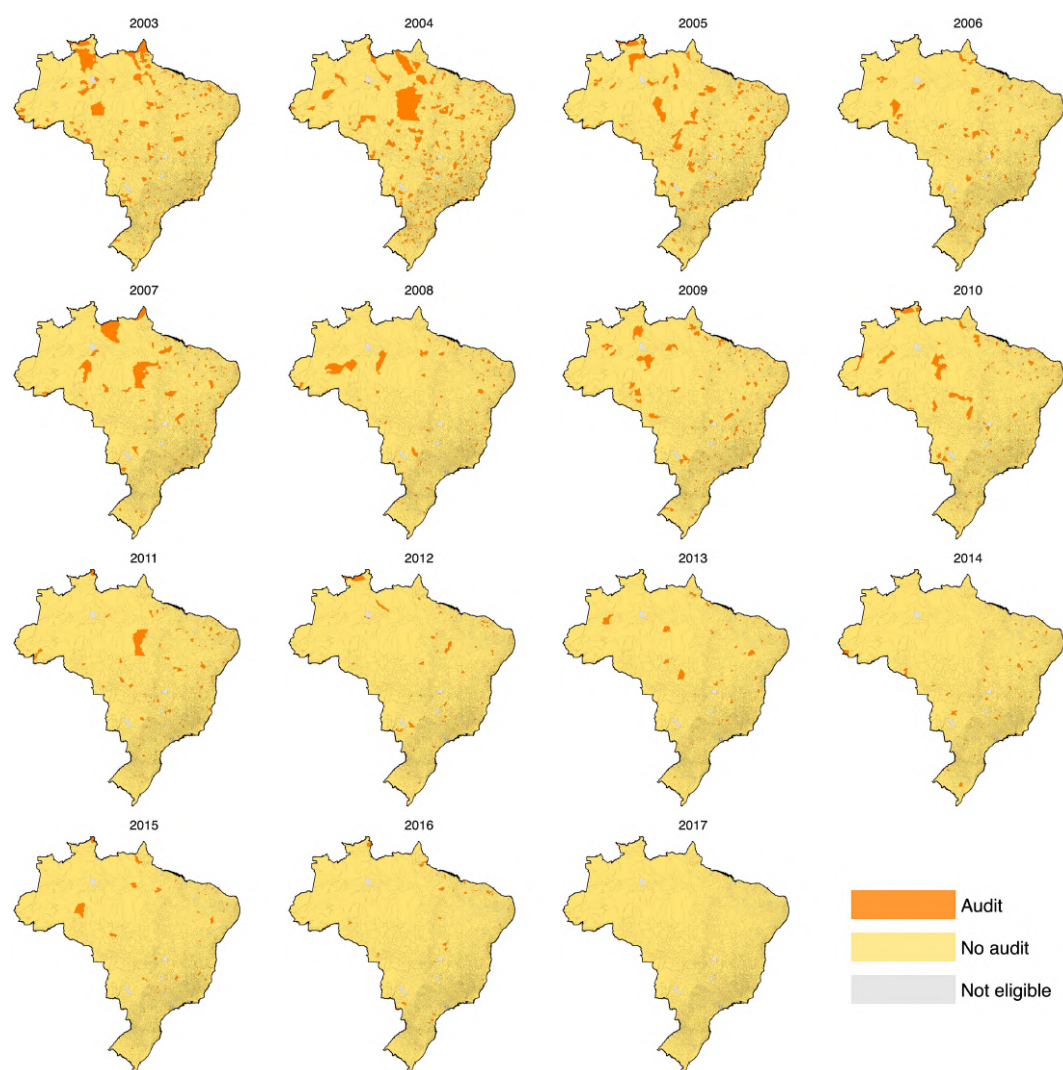


FIGURE A3.9
Distribution of Random Audits

Notes. This figure displays the geographic distribution of the random audits in each year under the Programa de Fiscalização em Entes Federativos (2003-2015) and the random third cycle of its successor, the Programa de Fiscalização em Entes Federativos (2016).

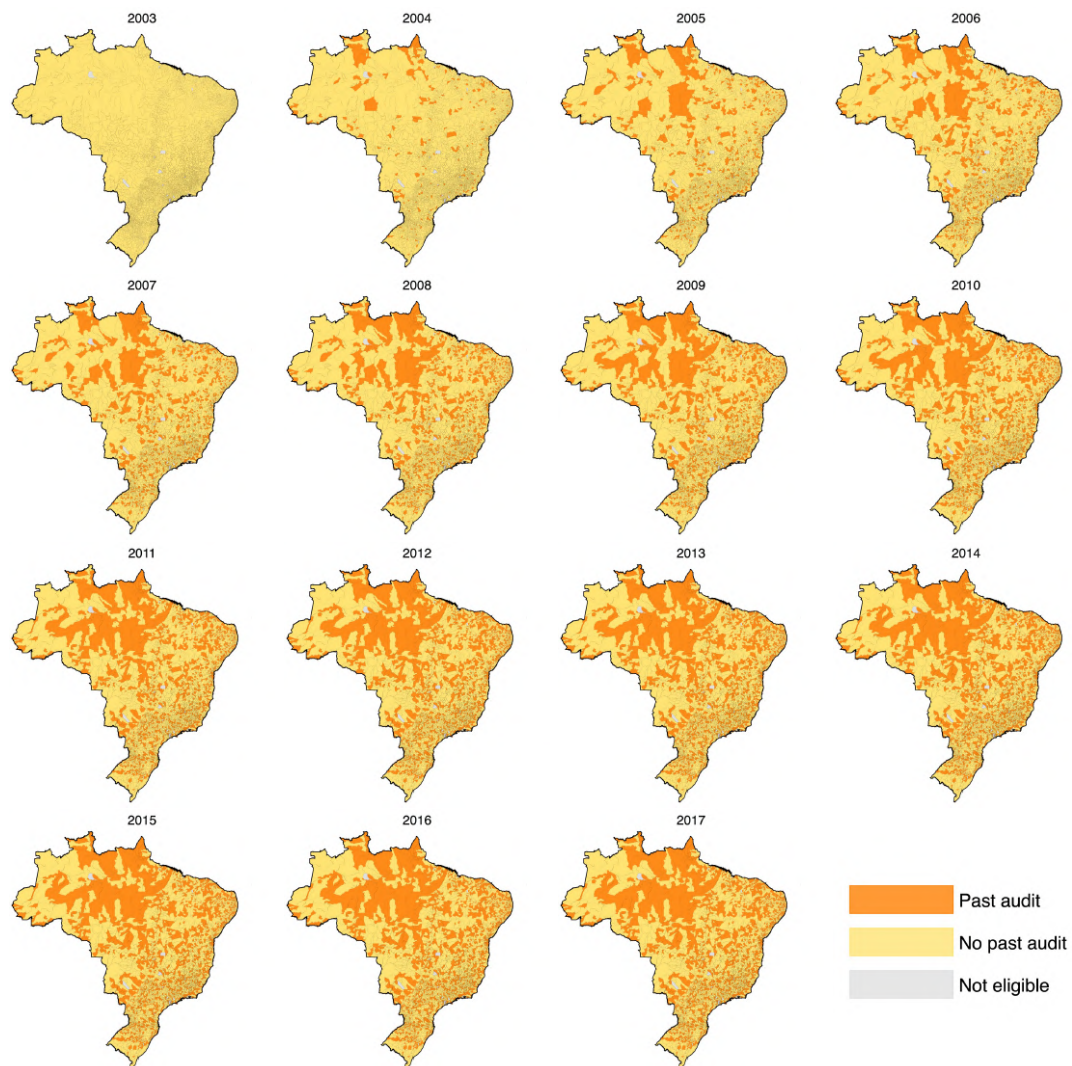


FIGURE A3.10
Distribution of the “Past Audit” Indicator

Notes. This figure displays the geographic distribution of the *Past audit* indicator for each year. The indicator takes value 1 if a municipality has been audited at random in a previous year and 0 otherwise.

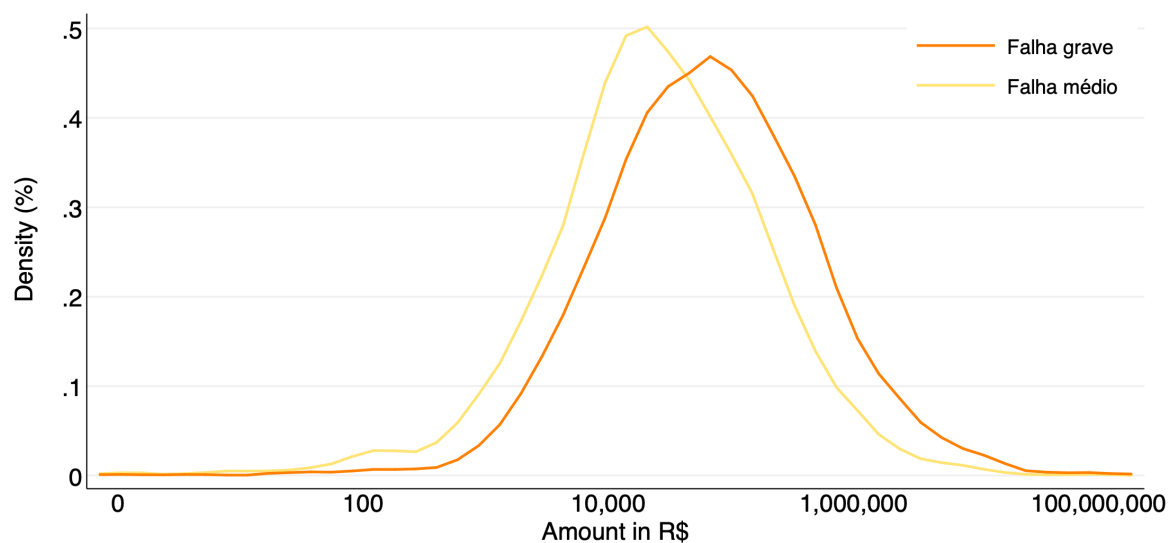


FIGURE A3.11

Financial Loss Uncovered by the Random Audits

Notes. This figure displays the considerable overlap between the financial losses judged as *falha média* or *falha grave*.

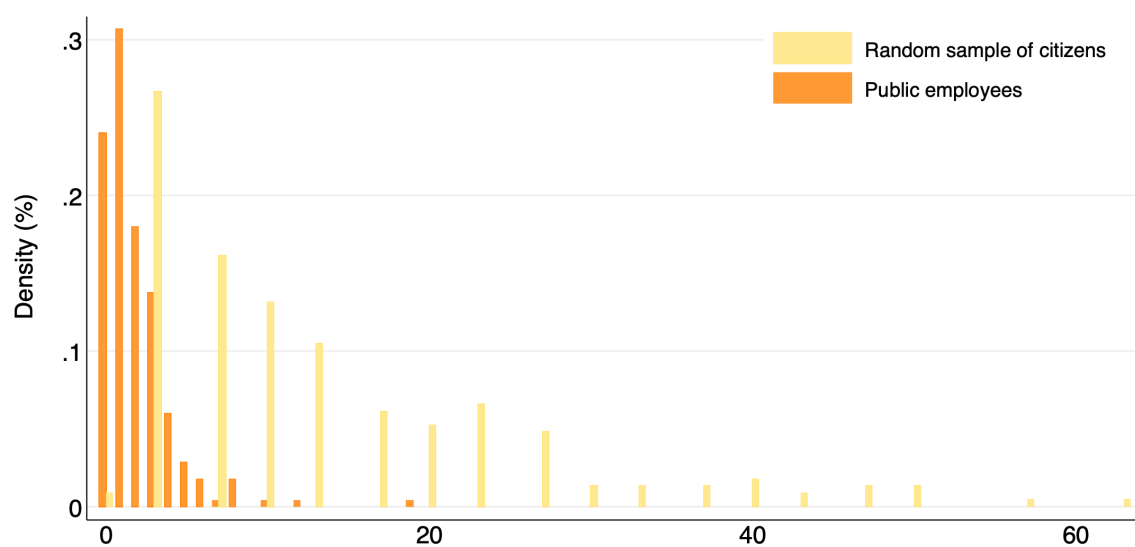


FIGURE A3.12

Rates of Illegitimate Payments for Citizens and Public Servants

Notes. This figure displays the distribution of the rate of illegitimate payments in municipalities uncovered by the random audits. The rate for citizens is estimated based on families that were randomly sampled ($N = 30$). The rate for public servants is estimated based on the auditor's cross-referencing of public employment records with Bolsa Família payments.

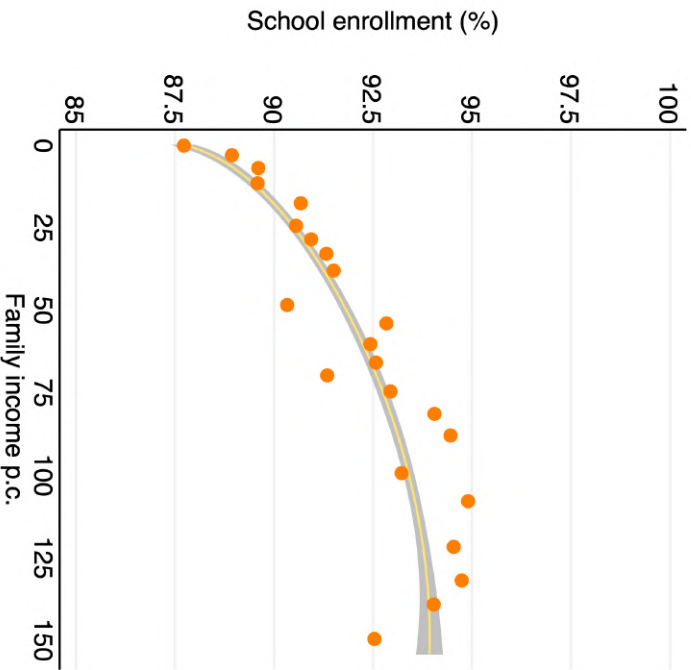


FIGURE A3.13

School Enrollment and Family Income

Notes. This figure displays the relationship between families' per capita income and children's school enrollment. The curve shows the fractional polynomial fit and 95% confidence interval for newly registered families who don't benefit from Bolsa Família. The dots represent the mean income and school enrollment for 30 equally sized bins. Standard errors are clustered at the family level.

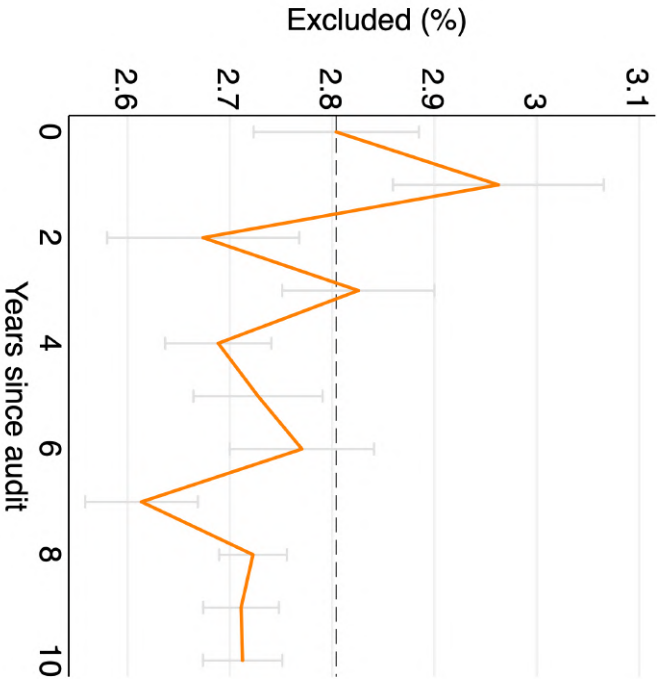


FIGURE A3.14

Change in Exclusions after a Random Audit

Notes. This figure displays the percentage of families that are sanctioned and excluded from Bolsa Família in the years after a municipality has been audited at random. Numbers are corrected for municipality and year fixed effects. Error bars indicate standard errors.

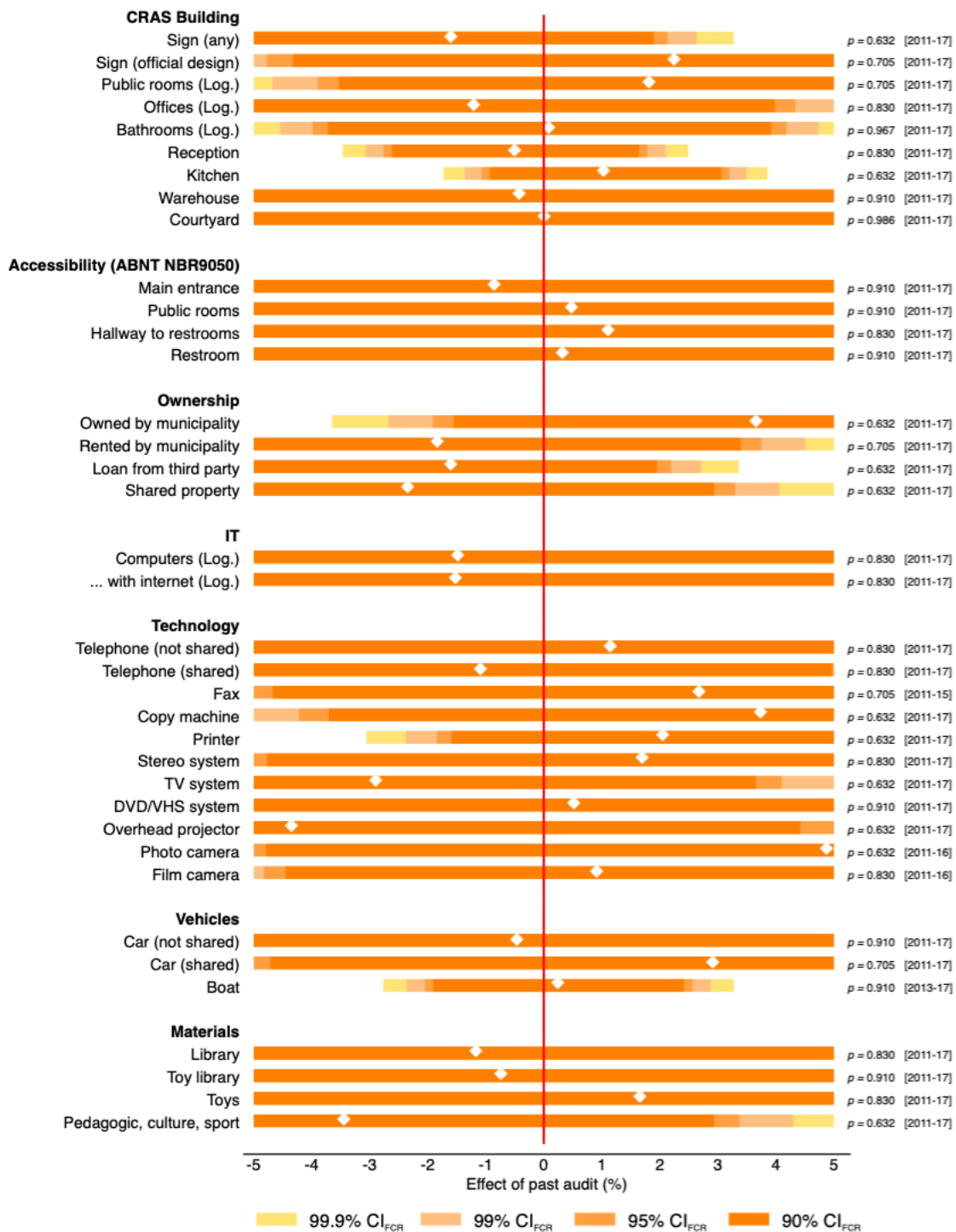


FIGURE A3.15

CRAS Infrastructure Doesn't Change after a Random Audit

Notes. This figure displays the change in the available infrastructure at Bolsa Família registration centers after a municipality has been audited at random. The white diamonds show the estimated treatment effect from 37 regressions where infrastructure variables in the Censo SUAS are regressed on the “Past audit” indicator, registration center and year fixed effects. Colors indicate false coverage rate adjusted 90%, 95%, 99%, and 99.9% confidence intervals (Benjamini and Yekutieli, 2005).

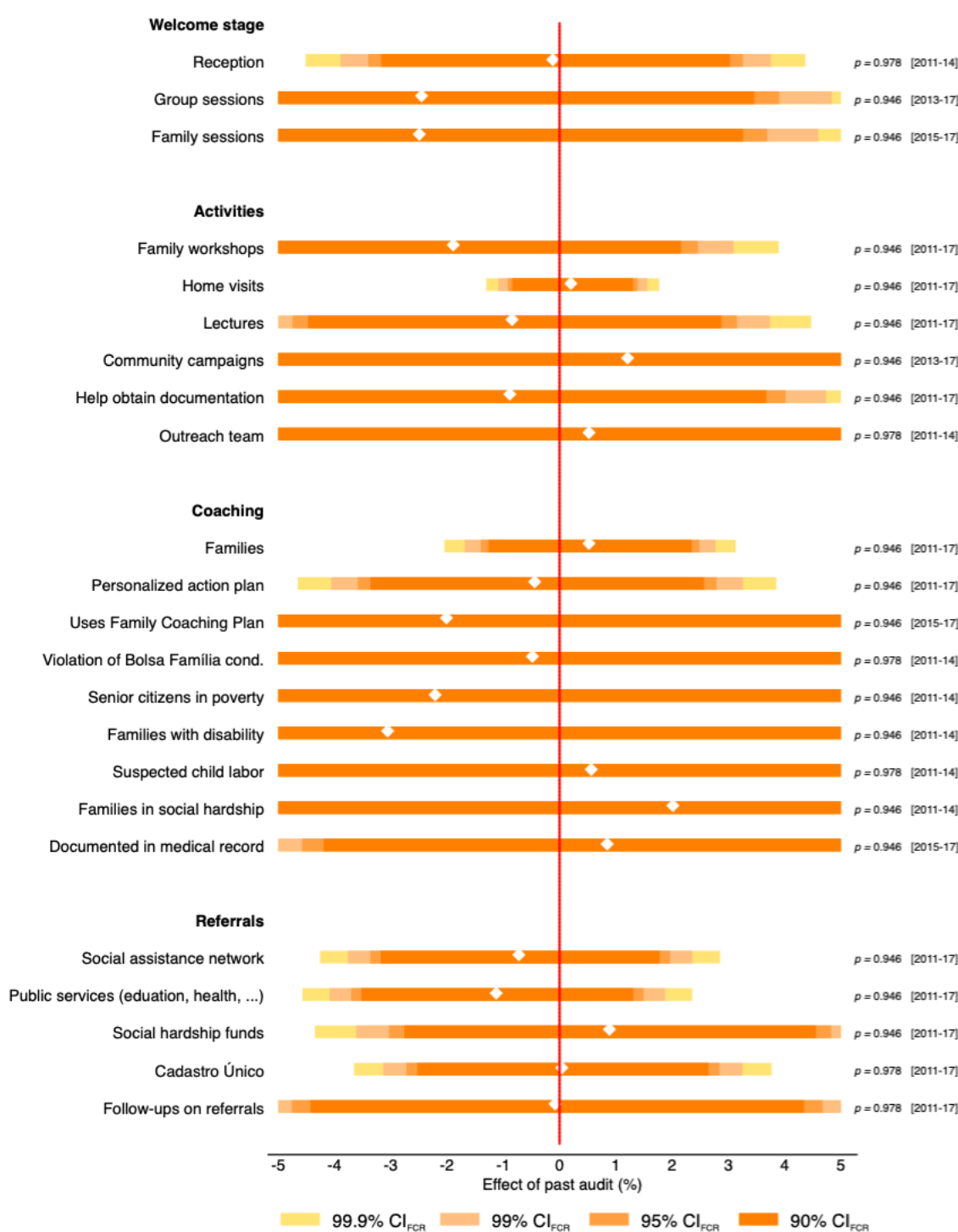


FIGURE A3.16

Complementary Actions Don't Change after a Random Audit

Notes. This figure displays the change in activities and programs aimed at vulnerable families through the Serviço de Proteção e Atendimento Integral à Família (PAIF) after a municipality has been audited at random. The white diamonds show the estimated treatment effect from 23 regressions where variables related to complementary programs in the Censo SUAS are regressed on the “Past audit” indicator, registration center and year fixed effects. Colors indicate false coverage rate adjusted 90%, 95%, 99%, and 99.9% confidence intervals (Benjamini and Yekutieli, 2005).

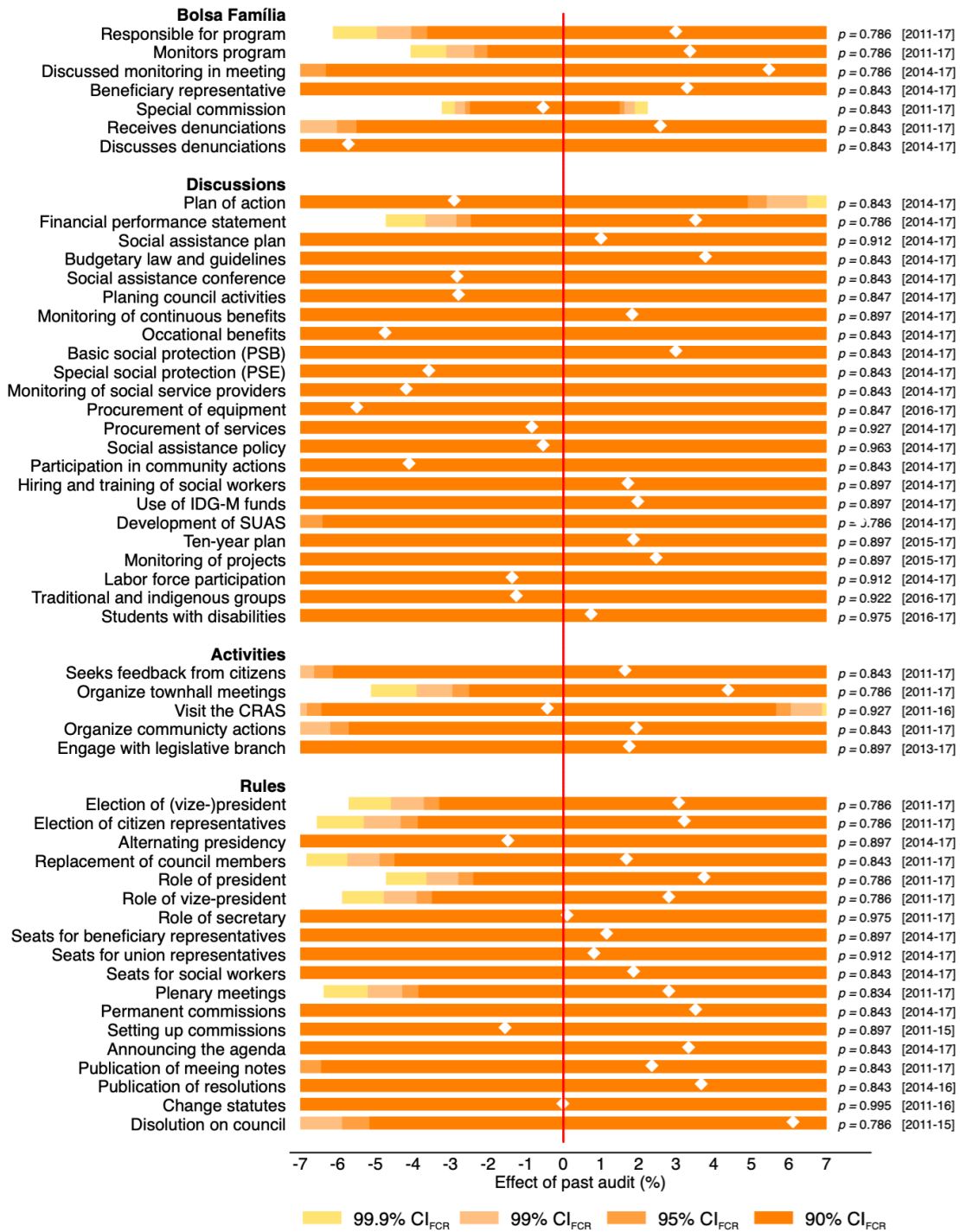


FIGURE A3.17

Governance of Social Programs Doesn't Change after a Random Audit

Notes. This figure displays the change in the governance of social programs after a municipality has been audited at random. The white diamonds show the estimated treatment effect from 53 regressions where variables related to municipalities' Councils for Social Assistance (CMAS) in the Censo SUAS are regressed on a "Past audit" indicator, municipality and year fixed effects. Colors indicate false coverage rate adjusted 90%, 95%, 99%, and 99.9% confidence intervals (Benjamini and Yekutieli, 2005).

VI. ADDITIONAL TABLES

TABLE A3.17
BOLSA FAMÍLIA INCREASES SCHOOL ENROLLMENT
(NO CHILD FIXED EFFECTS)

	(1) School enrollment (%) (All marginal families)	(2) School enrollment(%) (Newly registered families)	(3) School enrollment (%) (Same registration month)
BF	4.432*** (0.093)	3.104*** (0.145)	5.163*** (0.125)
Control mean	87.283	86.442	86.809
Child FE	No	No	No
Municipality FE	Yes	Yes	Yes
Year \times strata FE	Yes	Yes	Yes
R2	0.075	0.107	0.082
N(municipalities)	5,401	5,068	4,858
N(priority strata)	12,559	8,641	6,008
N(children)	2,573,117	590,630	747,786
N	5,146,234	1,181,260	1,495,572
Years	2012-2017	2012-2017	2012-2017

Notes. This table reports the effect of inclusion into the Bolsa Família program and random audits on children's school enrollment in a model without child fixed effects. The dependent variable in all models takes on value 100 if a child is enrolled in school and 0 otherwise. "BF" indicates if a child's family is included in the Bolsa Família program, "Past audit" indicates that a municipality has been audited at random, and "BF \times Past audit" is the interaction of the two treatments. Columns (1) and (2) present the results for the most representative sample. Columns (3) and (4) consider only newly registered families. Columns (5) and (6) require families to have last updated their data in the same month. All models include municipality and "Year \times Priority strata" fixed effects, but no individual-level child fixed effects. Standard errors are clustered at both the family and the municipality level. Significance levels: $^+ P < 0.1$, $^* P < 0.05$, $^{**} P < 0.01$, $^{***} P < 0.001$.

TABLE A3.18
CORRUPTION IN BOLSA FAMÍLIA AND THE EDUCATIONAL SYSTEM IS CORRELATED WITH
CORRUPTION IN OTHER SECTORS

	Bolsa Família			Education		
	(1) Irregularities (Log.)	(2) Mismanagement (Log.)	(3) Corruption (Log.)	(4) Irregularities (Log.)	(5) Mismanagement (Log.)	(6) Corruption (Log.)
Other corruption (Log.)	0.149*** (0.031)	-0.012 (0.025)	0.159*** (0.032)	0.805*** (0.041)	-0.006 (0.037)	0.909*** (0.040)
Other mismanagement (Log.)	-0.009 (0.017)	0.105*** (0.015)	-0.040* (0.018)	0.081*** (0.018)	0.553*** (0.021)	-0.031 (0.019)
Constant	1.712*** (0.165)	0.428 (0.308)	1.577*** (0.172)	-1.071*** (0.241)	-0.447* (0.226)	-1.388*** (0.237)
Inspection orders	Nonpar.	Nonpar.	Nonpar.	Nonpar.	Nonpar.	Nonpar.
State FE	Yes	Yes	Yes	Yes	Yes	Yes
Lottery FE	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.314	0.166	0.306	0.683	0.591	0.698
N	1134	1134	1134	1102	1102	1102

Notes. This table reports the relationship between corruption in the Bolsa Família program and the educational system with corruption in other parts of municipal government. The dependent variable in Columns (1) and (4) is the logarithm of the total number of irregularities uncovered by the random audit program. Columns (2) and (5) include only instances of mismanagement (*falha formal*), and Columns (4) and (6) only instances of corruption (*falha média* or *falha grave*). All models include the regressors of interest—the logarithms of the number of instances of corruption and mismanagement in other sectors—, fixed effects for the number of inspection orders, the state, and the round of the audit lottery. Robust standard errors are reported in brackets. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A3.19
BOLSA FAMÍLIA IS EQUALLY EFFECTIVE AFTER A RANDOM AUDIT FOR FAMILIES
REGISTERED AT HOME

	School enrollment (%) (All marginal families)		School enrollment(%) (Newly registered families)		School enrollment (%) (Same registration month)	
	(1) Home	(2) CRAS	(3) Home	(4) CRAS	(5) Home	(6) CRAS
BF	0.679*** (0.155)	0.900*** (0.072)	0.065 (0.342)	1.389*** (0.154)	0.834*** (0.192)	0.816*** (0.092)
Past audit	2.482 ⁺ (1.488)	-0.315 ⁺ (0.178)	3.555 ⁺ (1.937)	-0.482 ⁺ (0.292)	5.348*** (1.118)	-0.237 (0.322)
BF × Past audit	-0.078 (0.275)	0.302* (0.135)	0.278 (0.421)	0.518* (0.248)	-0.118 (0.367)	0.283 ⁺ (0.153)
Control mean	88.213	87.163	86.727	86.388	86.877	86.748
Child FE	Yes	Yes	Yes	Yes	Yes	Yes
Year × strata FE	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.936	0.927	0.952	0.933	0.946	0.933
N(municipalities)	3,601	5,382	2,194	4,979	2,284	4,802
N(priority strata)	5,157	12,350	3,066	8,376	1,967	5,881
N(children)	174,384	2,350,705	48,316	531,521	49,350	684,731
N	348,768	4,701,410	96,632	1,063,042	98,700	1,369,462
Years	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017	2012-2017

Notes. This table reports the effect of inclusion into the Bolsa Família program and random audits on children's school enrollment separately for families that registered during a home visit (Columns 1, 3, and 5) and those that registered at the CRAS (Columns 2, 4, and 6). The dependent variable in all models takes on value 100 if a child is enrolled in school and 0 otherwise. "BF" indicates if a child's family is included in the Bolsa Família program, "Past audit" indicates that a municipality has been audited at random, and "BF × Past audit" is the interaction of the two treatments. Columns (1) and (2) present the results for the most representative sample. Columns (3) and (4) consider only newly registered families. Columns (5) and (6) require families to have last updated their data in the same month. All models include individual-level child fixed effects, municipality fixed effects and "Year × Priority strata" fixed effects. Standard errors are clustered at both the family and the municipality level. Significance levels: ⁺ $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A3.20
CHANGE IN SANCTIONS AFTER A RANDOM AUDIT

	Log. excluded families		Log. withheld benefits	
	(1)	(2)	(3)	(4)
Past audit	-0.017 (0.042)		0.021 (0.021)	
Immediately after		0.080 (0.049)		0.049* (0.023)
Long run		-0.089* (0.045)		0.005 (0.023)
Control mean	3.311	3.311	5.085	5.085
Municipality FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
R2	0.855	0.855	0.933	0.933
N(municipalities)	5539	5539	5539	5539
N	33226	33226	38760	38760
Years	2012-2017	2012-2017	2011-2017	2011-2017

Notes. This table reports the effect of a random audit on the number of beneficiary families that are sanctioned. The dependent variable in Columns (1) and (2) is the logarithm of the number of families that are excluded from Bolsa Família. The dependent variable in Columns (3) and (4) is the logarithm of the number of families whose benefits are withheld for at least a month. “Past audit” indicates that a municipality has been audited at random. “Immediately after” takes value 1 if the municipality has been randomly audited in the previous year. “Long run” takes value 1 if the municipality has been randomly audited in an earlier year. All models include municipality and year fixed effects. Standard errors are clustered at the municipality level. Significance levels: $^+P < 0.1$, $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

TABLE A3.21
THE BOLSA FAMÍLIA MANAGEMENT INDEX DOESN'T CHANGE...

	(1)	(2)	(3)	(4)	(5)
	IGD-M	TAFE	TAAS	TAC	TCQC
	(Attendance monitoring)	(Medical check-ups)	(Data updating)	(CadU coverage)	
Past audit	-0.002 (0.008)	-0.004 ⁺ (0.003)	-0.004 (0.006)	0.002 (0.003)	-0.001 (0.005)
Control mean	.782	.877	.716	.759	.866
Municipality FE	Yes	Yes	Yes	Yes	Yes
Year × Month FE	Yes	Yes	Yes	Yes	Yes
R2	0.275	0.315	0.501	0.564	0.654
N	542630	841405	841456	841466	619906
Years	2010-2018	2006-2018	2006-2018	2006-2018	2006-2015

Notes. This table reports the effect of a random audit on the IGD-M (Column 1), the index of municipal management quality, and its subindices for school monitoring (Column 2), health monitoring (Column 3), the data actualization (Column 4), and Cadastro Único coverage (Column 5). “Past audit” indicates that a municipality has been audited at random. All models include municipality and Year × Month fixed effects. Standard errors are clustered at the municipality level. Significance levels: ⁺ $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A3.22
... BUT THE COMPONENTS DO.

	TAFE		TAAS		TAC		TCQC	
	(1) Children monitored (Log.)	(2) Children of beneficiaries (Log.)	(3) Families monitored (Log.)	(4) Families with check-ups (Log.)	(5) Updated entries (Log.)	(6) Registered families (Log.)	(7) Registered families (Log.)	(8) Eligible families (Log.)
Past audit	-0.023 ⁺ (0.014)	-0.025* (0.012)	-0.033 ⁺ (0.019)	-0.023 ⁺ (0.014)	-0.011 (0.013)	-0.012 (0.011)	-0.022** (0.008)	0.003 (0.007)
Control mean	6.80	6.91	6.34	6.57	7.03	7.34	7.37	7.22
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year × Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.970	0.992	0.955	0.987	0.987	0.991	0.998	0.995
N	437471	437471	437471	437471	437471	437471	215911	215911
Years	2012-2018	2012-2018	2012-2018	2012-2018	2012-2018	2012-2018	2012-2015	2012-2015

Notes. This table reports the effect of a random audit on the individual components of the IGD-M subindices: the logarithms of the number of children whose school attendance is monitored (Column 1) and the number of beneficiary children (Column 2), the number of families in the health monitoring system (Column 3) and the number of families subject to health conditionalities (Column 4), the number of entries in the Cadastro Único that are up to date (Column 5) and the number of families that are registered (Column 6), and the number of registered families (Column 7) and the estimated number of eligible families (Column 8). “Past audit” indicates that a municipality has been audited at random. All models include municipality and Year × Month fixed effects. Standard errors are clustered at the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A3.23
CRAS CENTERS DON'T UPDATE THE CADASTRO ÚNICO DIFFERENTLY AFTER A RANDOM AUDIT

	Center						Employees
	(1) Updates CadU (%)	(2) Special team (%)	(3) Only paper (%)	(4) Mostly paper (%)	(5) Mostly digital (%)	(6) Only digital (%)	(7) Updates CadU (%)
Past audit	0.125 (2.277)	-2.065 (2.426)	-23.543 (16.507)	9.334 (10.738)	7.313 (10.983)	6.896 (12.531)	1.333 (1.544)
Control mean	63.596	41.373	29.351	1.812	11.464	57.374	11.757
Center FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.573	0.525	0.697	0.511	0.566	0.675	0.157
N(centers)	7768	7761	4332	4332	4332	4332	7817
N	51957	44597	8664	8664	8664	8664	324993
Years	2011-17	2011-17	2016-17	2016-17	2016-17	2016-17	2014-17

Notes. This table reports the effect of a random audit on the way that Bolsa Familia registration centers (CRAS) update the Cadastro Único. The dependent variable in Column (1) is an indicator of whether the center updates the Cadastro Único. Column (2) reports whether the center has a special team to do so. Columns (3) to (6) indicate whether the center initially collects information on paper that is later digitized or whether information is collected digitally. The dependent variable in Column (7) is an indicator of whether a given employee updates the Cadastro Único. The dependent variables in all models are scaled to take value 100 if the condition is met and 0 otherwise. “Past audit” indicates that a municipality has been audited at random. All models include registration center and year fixed effects. Standard errors are clustered at the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A3.24
CRAS CENTERS DON'T EMPLOY DIFFERENT PEOPLE AFTER A RANDOM AUDIT

Demographics			Education			Experience			Position		
(1) Male (%)	(2) Age (Years)	(3) Primary or less (%)	(4) High school (%)	(5) Some college (%)	(6) College (%)	(7) Postgrad (%)	(8) Tenure (Years)	(9) Hours (Weekly)	(10) CLT (%)	(11) Temporary Art. 37 (%)	(12) Commission Art. 37 (%)
Past audit	0.058 (0.688)	-0.063 (0.213)	0.196 (0.994)	0.840 (0.760)	-1.179 (1.102)	0.485 (0.868)	0.147 (0.159)	-0.138 (0.364)	0.072 (0.908)	0.090 (2.165)	-0.850 (0.762)
Control mean	16.736	36.812	7.831	9.779	41.283	6.100	2.864	34.651	7.027	33.892	9.917
Center FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.069	0.121	0.073	0.092	0.080	0.109	0.204	0.217	0.423	0.381	0.200
N(centers)	7868	7790	7868	7868	7868	7868	7793	7868	7868	7868	7868
N	465883	326605	465883	465883	465883	465883	251436	465883	465883	465883	465883
Years	2012-17	2012-16	2012-17	2012-17	2012-17	2012-17	2015-17	2012-17	2012-17	2012-17	2012-17

Notes. This table reports the effect of a random audit on the workforce at Bolsa Família registration centers (CRAS). The dependent variable in Column (1) takes value 100 if an employee is male. The dependent variable in Column (2) is the employee's age in years. Columns (3) to (7) are binary indicators for an employee's educational level. The dependent variable in Column (8) is an employee's tenure in years and Column (9) the weekly working hours, based on a categorical variable. The dependent variables in Columns (10), (11), and (12) indicate whether an employee is hired under the relatively strict Consolidação das Leis do Trabalho, or temporarily appointed or commissioned under the discretionary Art. 37. All binary indicators are scaled to take value 100 if the condition is met and 0 otherwise. "Past audit" indicates that a municipality has been audited at random. All models include registration center and year fixed effects. Standard errors are clustered at the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A3.25
BENEFICIARY INVOLVEMENT AT CRAS DOESN'T CHANGE MUCH AFTER A RANDOM AUDIT

	Access	Beneficiary involvement				Beneficiary representation				
	(1) Open (Hours)	(2) None (%)	(3) Informal, occas. (%)	(4) Informal, regular (%)	(5) Formal (%)	(6) Invites public (%)	(7) Financial support (%)	(8) Citizen rep. (%)	(9) Rep. is elected (%)	(10) Committees (%)
Past audit	-0.142 (0.117)	7.590 (4.689)	-4.454 (5.793)	-2.454 (3.833)	-0.683 (1.850)	0.216 (3.228)	-1.200 (0.795)	0.288 (2.522)	-2.506+ (1.411)	-5.435+ (3.255)
Control mean	45.526	31.552	45.572	16.316	6.560	23.719	1.319	9.863	2.311	5.977
Center FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.464	0.504	0.424	0.443	0.450	0.466	0.351	0.427	0.378	0.390
N(centers)	7768	7708	7708	7708	7708	7708	7708	7708	7708	7708
N	51956	30431	30431	30431	30431	30431	30431	30431	30431	30431
Years	2011-17	2014-17	2014-17	2014-17	2014-17	2014-17	2014-17	2014-17	2014-17	2014-17

Notes. This table reports the effect of a random audit on efforts for beneficiary participation at Bolsa Familia registration centers (CRAS). The dependent variable in Column (1) is the number of hours the CRAS is open each week. Columns (2) to (4) indicate how often and in what form the CRAS involves beneficiaries in planning activities: never, occasionally and informally, informally but regularly, or through a formal and established process. The dependent variables in Columns (6) to (10) indicate what forms of beneficiary participation exist at the registration center: whether the center invites beneficiaries to planning meetings (Column 6) and whether it offers financial support for such meetings (Column 7), whether there is a beneficiary representative (Column 8) and whether she is elected (9), and whether the CRAS encourages the formation of beneficiary committees or collectives (Column 10). All binary indicators are scaled to take value 100 if the condition is met and 0 otherwise. "Past audit" indicates that a municipality has been audited at random. All models include registration center and year fixed effects. Standard errors are clustered at the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A3.26
MUNICIPALITIES DON'T SPEND MORE ON SOCIAL ASSISTANCE

	Total social assistance			Administration			Children			Senior citizens			Disability		
	(1) Per capita	(2) Absolute	(3) % total	(4) Per capita	(5) Absolute	(6) % total	(7) Per capita	(8) Absolute	(9) % total	(10) Per capita	(11) Absolute	(12) % total	(13) Per capita	(14) Absolute	(15) % total
Past audit	-0.041 ⁺ (0.022)	-0.038 ⁺ (0.022)	-0.002 (0.002)	-0.029 (0.192)	-0.022 (0.212)	-0.006 (0.007)	-0.072 (0.057)	-0.086 (0.088)	-0.000 (0.001)	0.021 (0.107)	-0.074 (0.200)	-0.000 (0.000)	-0.073 (0.120)	-0.203 (0.282)	-0.001 (0.000)
Control mean	4.547	13.917	0.046	3.359	12.891	0.018	2.559	11.783	0.008	1.382	10.119	0.002	1.163	10.077	0.002
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.887	0.942	0.756	0.944	0.950	0.906	0.779	0.754	0.269	0.776	0.772	0.976	0.876	0.859	0.850
N	26545	26545	26530	3946	3946	3946	23286	23286	23260	12673	12673	12661	5550	5550	5546
Years	2013-17	2013-17	2013-17	2016-17	2016-17	2016-17	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17

Notes. This table reports the effect of a random audit on municipalities' social assistance expenditure. Columns (1) to (3) present the effect on total social assistance expenditure—in per capita terms, on the absolute amount, and as a percentage of total municipal expenditure. Columns (4) to (6) present the same metrics for expenditure on the administration of social programs, Columns (7) to (9) for expenditure on social assistance programs focused on children, Columns (10) to (12) for expenditure on social assistance for the elderly, and Columns (13) to (15) for expenditure on disability-related social assistance. "Past audit" indicates that a municipality has been audited at random. All models include municipality and year fixed effects. Standard errors are clustered at the municipality level. Significance levels: ⁺ $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A3.27
MUNICIPALITIES DON'T SPEND MORE ON EDUCATION

	Total educ. expenditure			Elementary schools			Middle school			Vocational schools			Higher education		
	(1) Per capita	(2) Absolute	(3) % total	(4) Per capita	(5) Absolute	(6) % total	(7) Per capita	(8) Absolute	(9) % total	(10) Per capita	(11) Absolute	(12) % total	(13) Per capita	(14) Absolute	(15) % total
Past audit	-0.007 (0.013)	-0.003 (0.012)	-0.004 (0.007)	-0.012 (0.018)	-0.008 (0.018)	0.027 (0.031)	0.120 (0.124)	0.119 (0.232)	0.004 (0.004)	0.048 (0.161)	0.000 (0.282)	0.000 (0.001)	-0.085 (0.091)	-0.184 (0.154)	0.000 (0.001)
Control mean	6.652	16.035	0.360	6.356	15.735	0.280	2.153	11.135	0.008	1.460	10.581	0.004	2.361	11.375	0.008
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.893	0.987	0.782	0.813	0.950	0.256	0.867	0.859	0.810	0.876	0.884	0.854	0.886	0.863	0.840
N	26584	26584	26569	26023	26023	25991	6587	6587	6581	2433	2433	2429	10270	10270	10262
Years	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17	2013-17

Notes: This table reports the effect of a random audit on municipalities' educational expenditure. Columns (1) to (3) present the effect on total educational expenditure—in per capita terms, on the absolute amount, and as a percentage of total municipal expenditure. Columns (4) to (6) present the same metrics for expenditure on elementary schools, Columns (7) to (9) for expenditure for middle schools, Columns (10) to (12) for expenditure on vocational training schools, and Columns (13) to (15) for expenditure for higher education. "Past audit" indicates that a municipality has been audited at random. All models include municipality and year fixed effects. Standard errors are clustered at the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A3.28
WHISTLEBLOWING DOESN'T CHANGE AFTER A RANDOM AUDIT

	Beneficiaries			Non-beneficiaries			Administrators			
	(1) Complaints about CRAS (Log.)	(2) Total denunciations (Log.)	(3) Illegitimate benefit (Log.)	(4) Retained benefit card (Log.)	(5) Total denunciations (Log.)	(6) Illegitimate benefit (Log.)	(7) Retained benefit card (Log.)	(8) Total denunciations (Log.)	(9) Illegitimate benefit (Log.)	(10) Retained benefit card (Log.)
Past audit	-0.005 (0.003)	-0.004 (0.006)	-0.005 (0.006)	0.001 (0.001)	0.016 (0.016)	0.019 (0.016)	0.000 (0.002)	-0.000 (0.002)	0.001 (0.002)	-0.000 (0.000)
Control mean	0.009	0.015	0.014	0.000	0.157	0.153	0.001	0.004	0.003	0.000
Municipality FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.173	0.206	0.209	0.125	0.449	0.449	0.126	0.141	0.139	0.125
N	44328	44328	44328	44328	44328	44328	44328	44328	44328	44328
Years	2010-17	2010-17	2010-17	2010-17	2010-17	2010-17	2010-17	2010-17	2010-17	2010-17

Notes. This table reports the effect of a random audit on the number of denunciations and complaints the MDS receives from the municipality. The dependent variable in Column (1) is the logarithm of the number of complaints about the CRAS or its workforce. Columns (2) to (4) present the effect on denunciations received from Bolsa Familia beneficiaries—on the logarithm of the total number of denunciations, the logarithm of the number of denunciations about illegitimate receipts of Bolsa Familia payments, and the logarithm of denunciations concerning retained benefit cards. Columns (5) to (7) present the same metrics for denunciations received from non-beneficiaries and Columns (8) to (10) for denunciations received from administrative staff. “Past audit” indicates that a municipality has been audited at random. All models include municipality and year fixed effects. Standard errors are clustered at the municipality level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

REFERENCES

- AVIS, E., C. FERRAZ, AND F. FINAN (2018): “Do Government Audits Reduce Corruption? Estimating the Impacts of Exposing Corrupt Politicians,” *Journal of Political Economics*, 126, 1912–1964.
- BENJAMINI, Y. AND Y. HOCHBERG (1995): “Controlling the false discovery rate: A practical and powerful approach to multiple testing,” *Journal of the Royal Statistical Society. Series B*, 57, 289–300.
- BENJAMINI, Y. AND D. YEKUTIELI (2005): “False discovery rate-adjusted multiple confidence intervals for selected parameters,” *Journal of the American Statistical Association*, 100, 71–81.
- BUCCIOL, A. AND M. PIOVESAN (2011): “Luck or cheating? A field experiment on honesty with children,” *Journal of Economic Psychology*, 32, 73–78.
- COHN, A., E. FEHR, AND M. A. MARÉCHAL (2014): “Business culture and dishonesty in the banking industry,” *Nature*, 516, 86.
- COHN, A., M. A. MARÉCHAL, AND T. NOLL (2015): “Bad boys: How criminal identity salience affects rule violation,” *The Review of Economic Studies*, 82, 1289–1308.
- FISCHBACHER, U. AND F. FÖLLMI-HEUSI (2013): “Lies in disguise – an experimental study on cheating,” *Journal of the European Economic Association*, 11, 525–547.
- HOUSER, D., S. VETTER, AND J. WINTER (2012): “Fairness and cheating,” *European Economic Review*, 56, 1645–1655.
- KRUPKA, E. L. AND R. A. WEBER (2013): “Identifying social norms using coordination games: Why does dictator game sharing vary?” *Journal of the European Economic Association*, 11, 495–524.
- PEI, Z., J.-S. PISCHKE, AND H. SCHWANDT (2019): “Poorly measured confounders are more useful on the left than on the right,” *Journal of Business & Economic Statistics*, 37, 205–216.

Appendix A4

Appendices to:

Genes, Pubs, and Drinks:

Gene-Environment Interplay and
Alcohol Licensing Policy in the UK

I. ADDITIONAL FIGURES

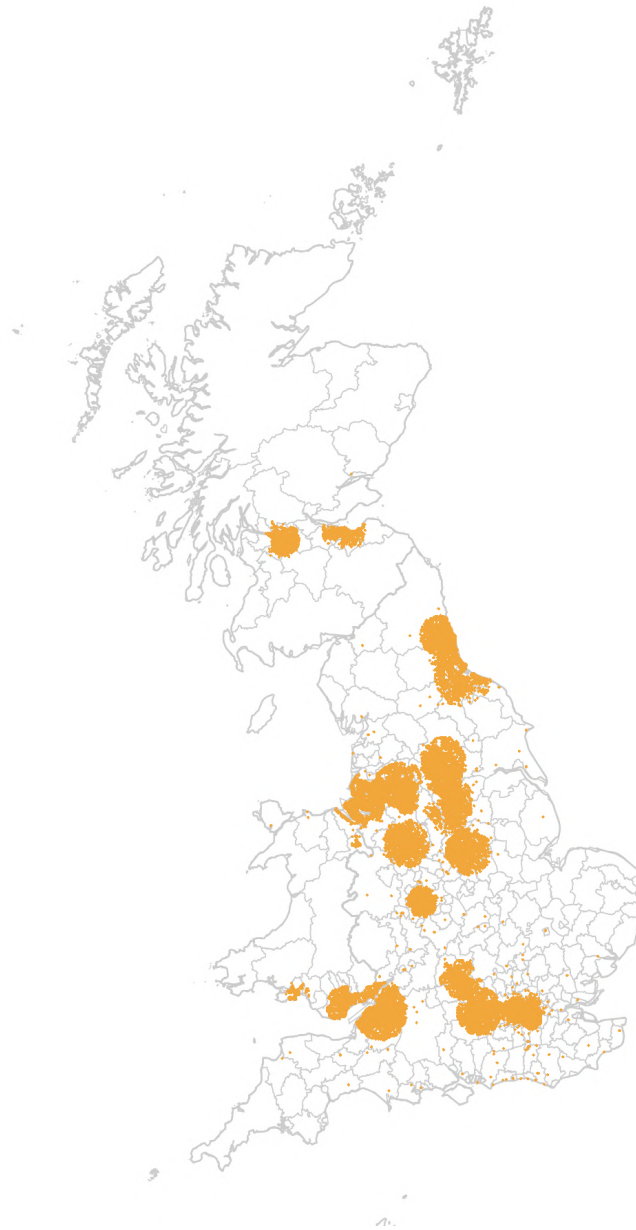


FIGURE A4.1

Geographic Distribution of UK Biobank Participants

Notes. This figure displays the geographic distribution of participants in the UK Biobank. Participants were recruited by one of the UKB's 23 assessment centers. The location of assessment centers was chosen such that at least 150,000 potential participants in this age group live within less than 10 miles of each center (UKB, 2006, p.49). As a result, the vast majority of participants live in urban areas.

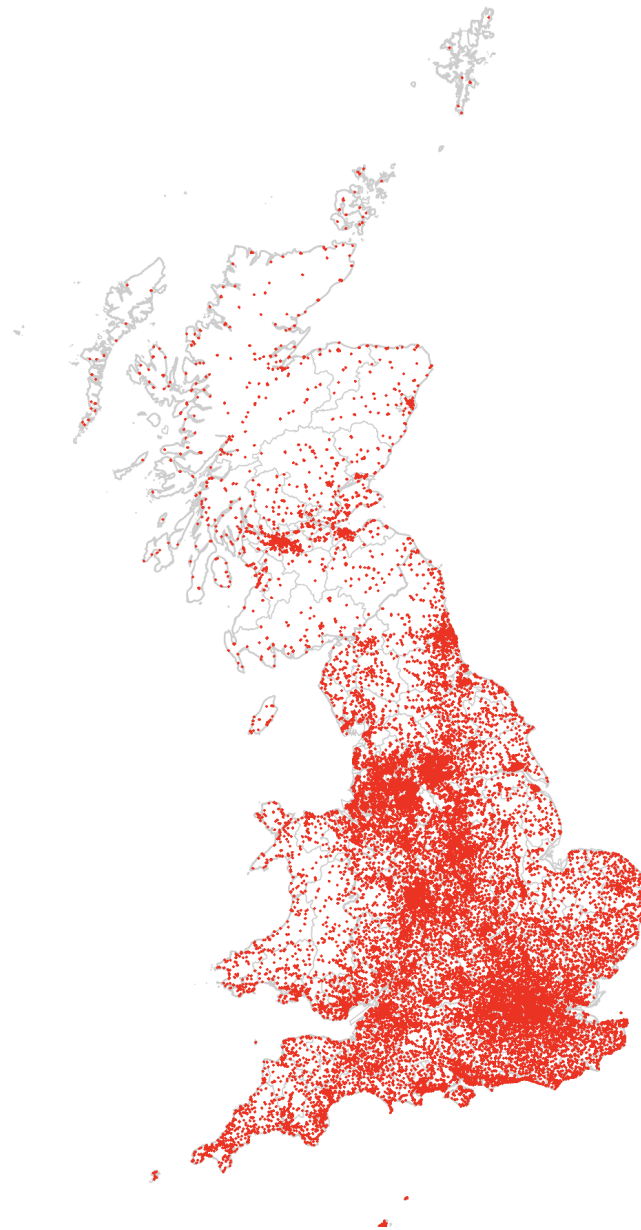


FIGURE A4.2

Geographic Distribution of Pubs in the United Kingdom

Notes. This figure displays the geographic distribution of pubs in the United Kingdom from The Good Pub Guide.

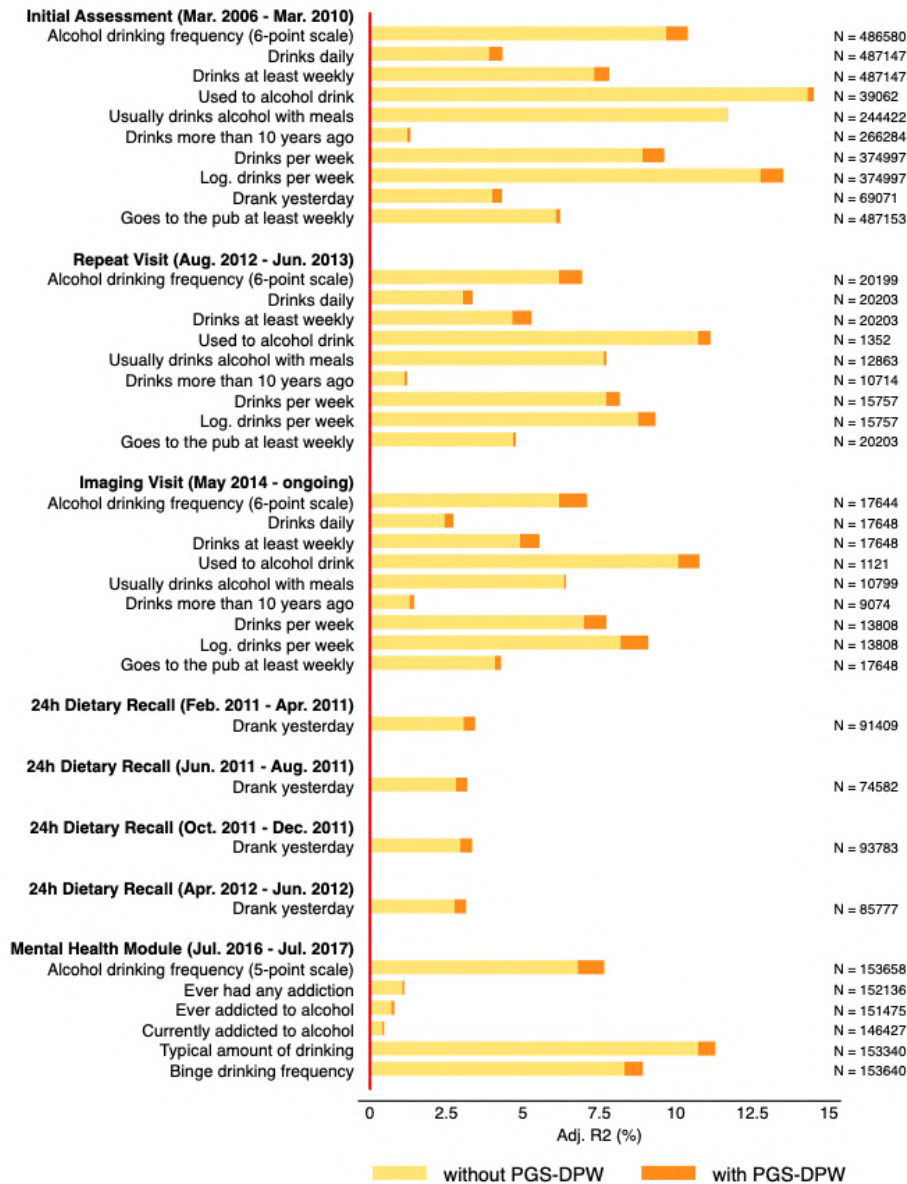


FIGURE A4.3
Predictive Power of the Polygenic Score

Notes. This figure shows the predictive power of the polygenic score for different drinking-related outcomes in the UKB. Yellow bars indicate the adjusted R^2 from 38 regressions where drinking-related outcomes in the UKB are regressed on the first 40 principal components of the genetic data, “Age \times Sex” fixed effects, and local authority fixed effects. Orange bars indicate the increase in adjusted R^2 when the polygenic score for the number of drinks per week is included in the regression.

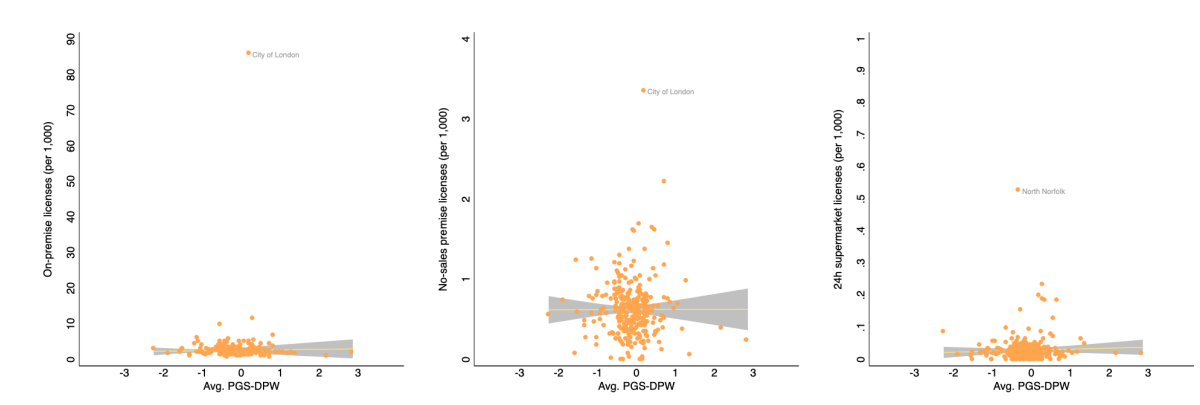


FIGURE A4.4

Local Licensing Policy Is Uncorrelated with the Average Genetic Predisposition

Notes. This figure shows the relationship between local licensing policy and the average genetic predisposition. The left panel shows the correlation with the number of premises per capita allowed to sell alcohol on-premise (including premises permitted to sell alcohol both on- and off-premise). The middle panel shows the correlation with the number of premises per capita allowed to provide licensable entertainment but not permitted to sell alcohol. The right panel shows the correlation with the number of supermarkets per capita licensed to sell alcohol 24h a day. The average genetic predisposition is measured as the mean polygenic score for the number of drinks per week for UKB participants living in the local authority. Orange points represent the values of different local authorities, with labeled outliers (values more than five standard deviations from the mean). The yellow lines and the shaded gray areas show the linear fit and the associated 95% confidence intervals.

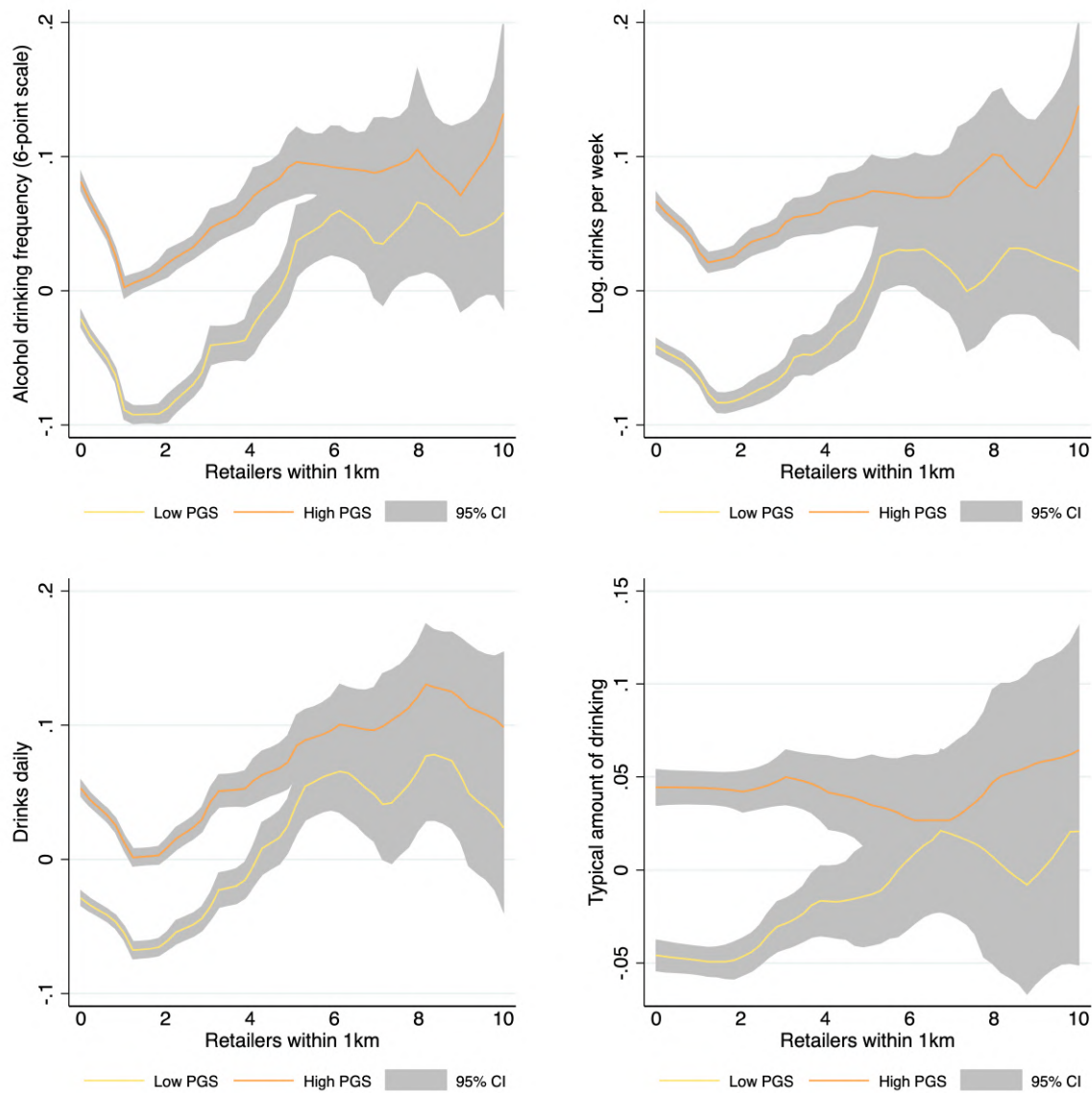


FIGURE A4.5
Participants with a High PGS Have Less Elastic Demand (Retailers)

Notes. This figure displays the relationship between drinking behavior and the number of retailers within 1000m of UKB participants with a polygenic score above and below the median. The orange and yellow lines and the shaded gray areas show the polynomial fit and the associated 95% confidence interval for UKB participants with high and low polygenic scores, respectively. The figure was constructed using kernel-weighted local polynomial smoothing. All models control for genetic stratification using the first 40 principal components of the genetic data, “Age \times Sex” fixed effects, and local authority fixed effects.

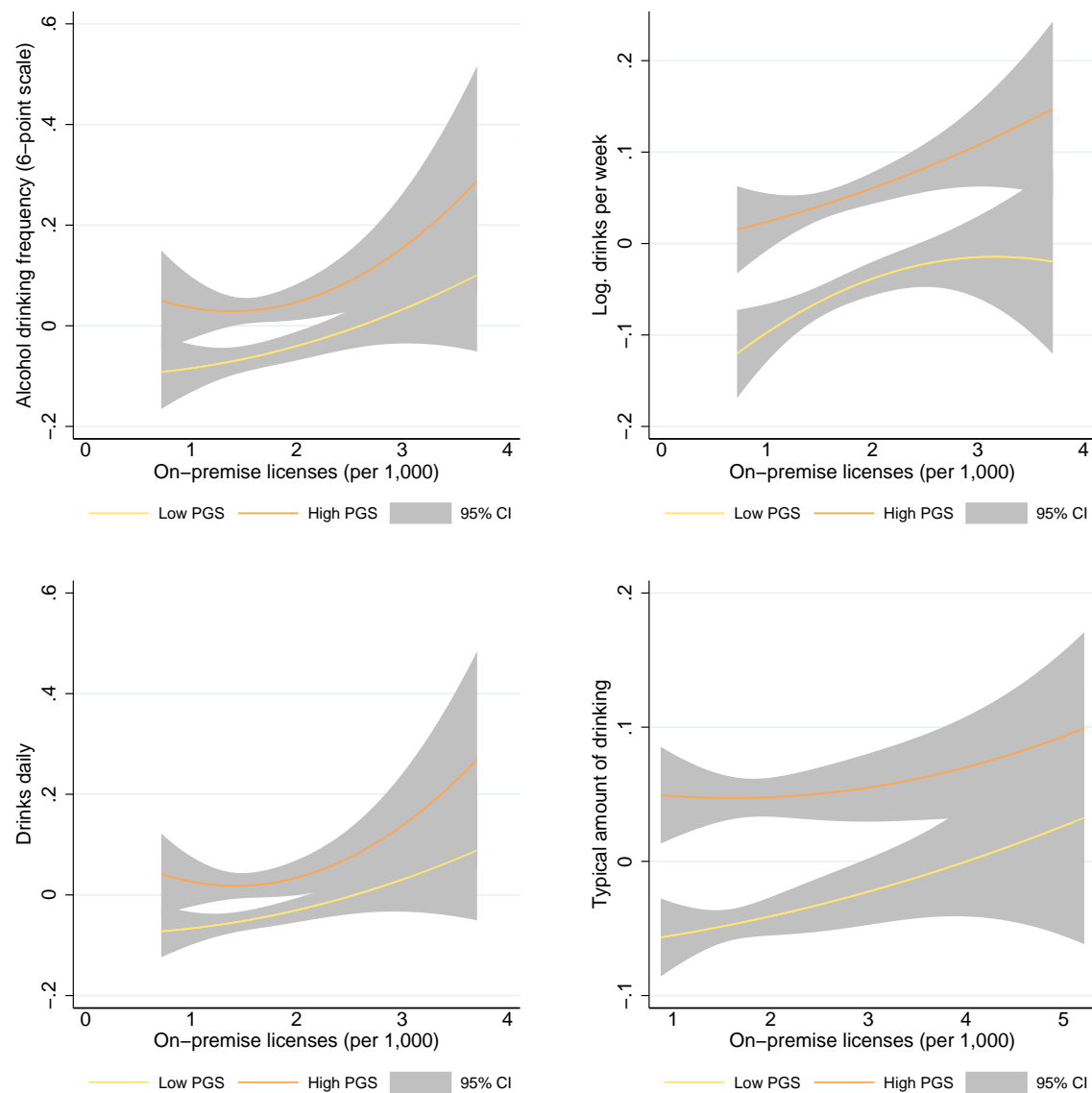


FIGURE A4.6
High PGS Participants React Less to Policy (Licensed Premises)

Notes. This figure displays the relationship between drinking behavior and the number of premises per capita allowed to sell alcohol on-premise (including premises permitted to sell alcohol both on- and off-premise) for UKB participants with a polygenic score above and below the median. The orange and yellow lines and the shaded gray areas show the polynomial fit and the associated 95% confidence interval for UKB participants with high and low polygenic scores, respectively. All models control for genetic stratification using the first 40 principal components of the genetic data, “Age \times Sex” fixed effects. Standard errors are clustered at the local authority level.

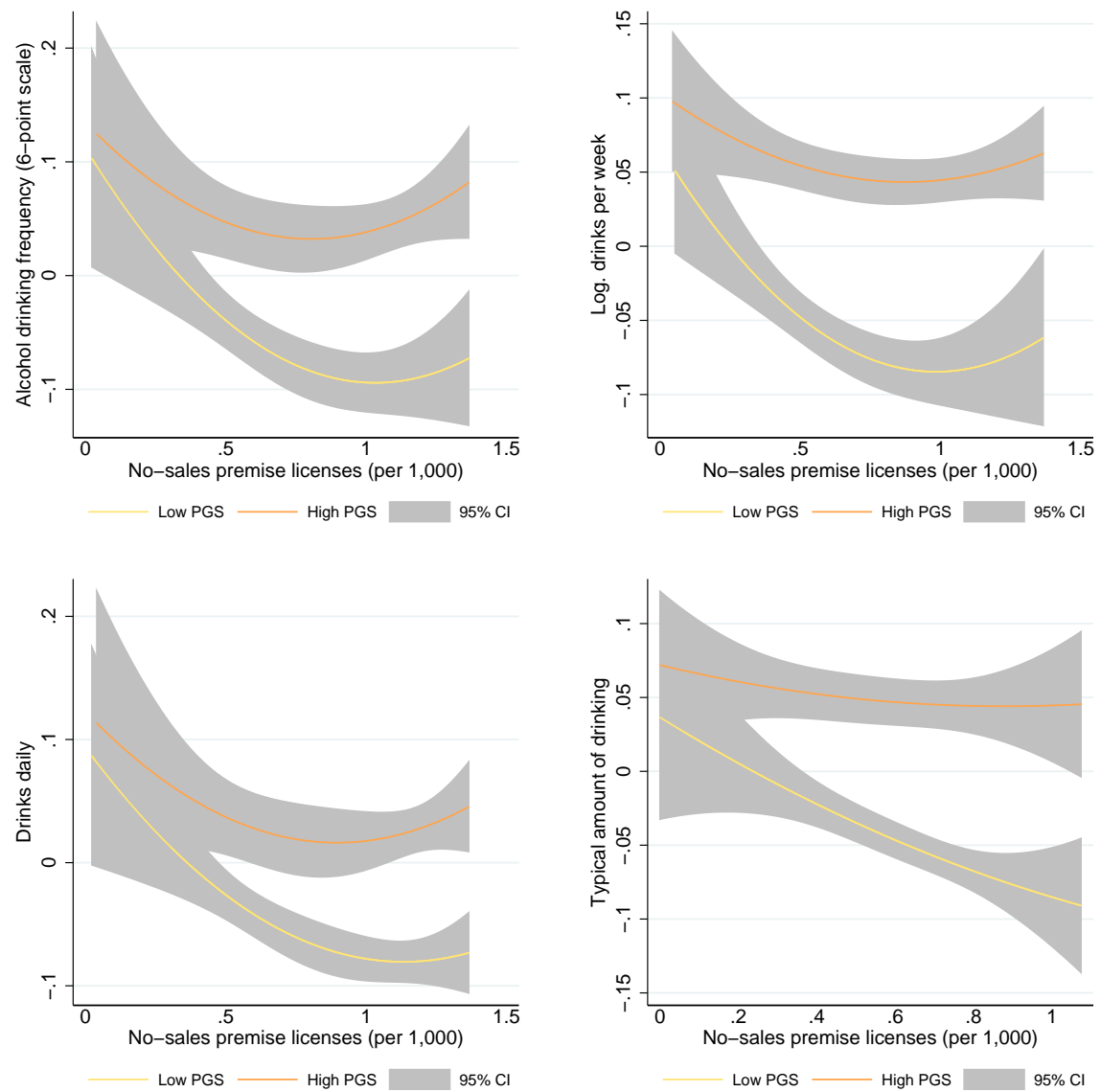


FIGURE A4.7

High PGS Participants React Less to Policy (No-Sale Licenses)

Notes. This figure displays the relationship between drinking behavior and the number of premises per capita allowed to provide licensable entertainment but not permitted to sell alcohol for UKB participants with a polygenic score above and below the median. The orange and yellow lines and the shaded gray areas show the polynomial fit and the associated 95% confidence interval for UKB participants with high and low polygenic scores, respectively. All models control for genetic stratification using the first 40 principal components of the genetic data, “Age \times Sex” fixed effects. Standard errors are clustered at the local authority level.

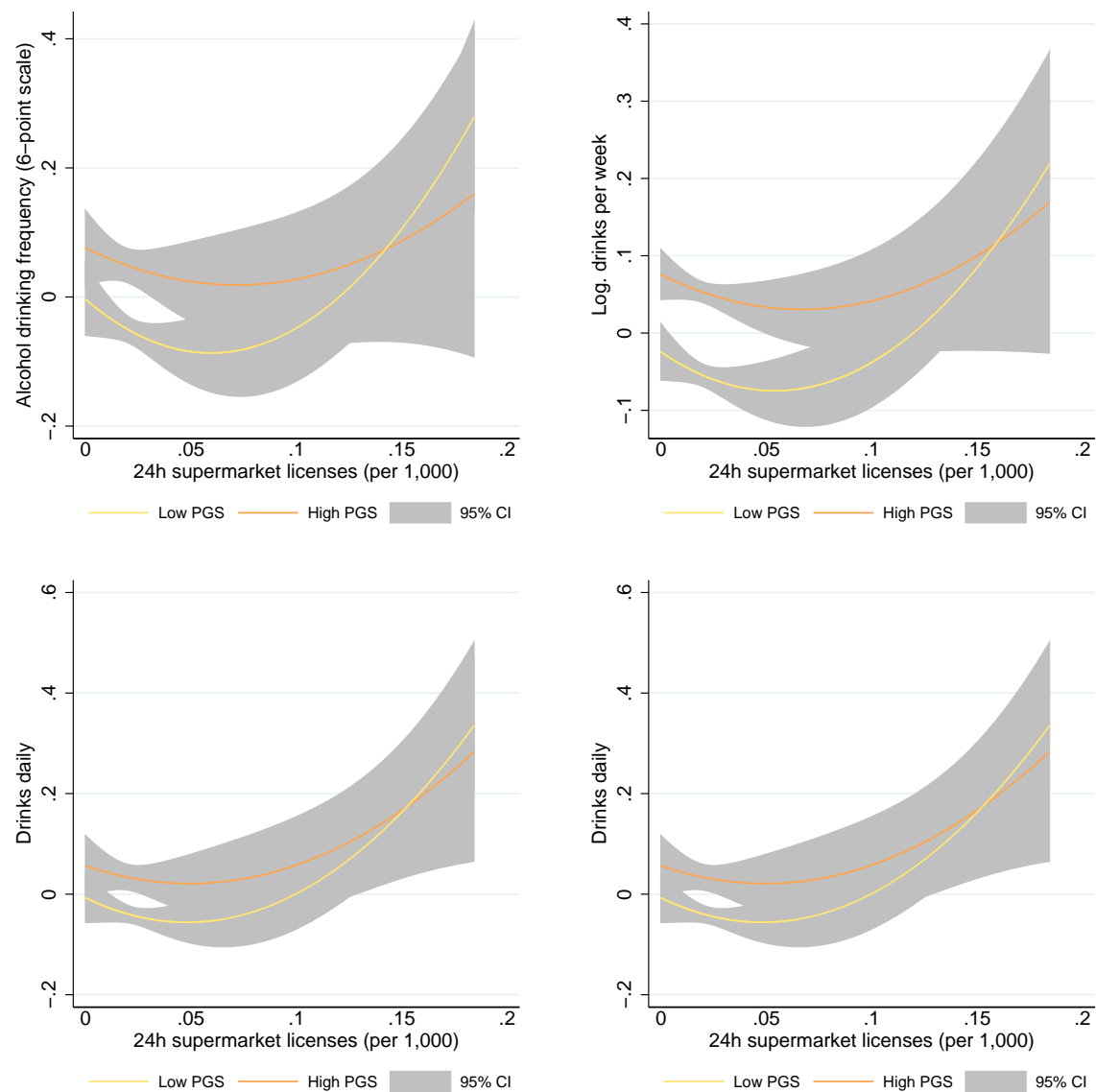


FIGURE A4.8

High PGS Participants React Less to Policy (24h Supermarkets)

Notes. This figure displays the relationship between drinking behavior and the number of supermarkets per capita licensed to sell alcohol 24h a day for UKB participants with a polygenic score above and below the median. The orange and yellow lines and the shaded gray areas show the polynomial fit and the associated 95% confidence interval for UKB participants with high and low polygenic scores, respectively. All models control for genetic stratification using the first 40 principal components of the genetic data, “Age \times Sex” fixed effects. Standard errors are clustered at the local authority level.

II. ADDITIONAL TABLES

TABLE A4.1
SUMMARY STATISTICS OF DRINKING OUTCOMES IN THE UKB

	count	mean	sd	min	max
Initial Assessment (Mar 2006-Mar 2010)					
Alcohol drinking frequency (6-point scale)	501088	3.050	1.531	0	5
Drinks daily	501693	0.204	0.403	0	1
Drinks at least weekly	501693	0.692	0.462	0	1
Used to alcohol drink	40503	0.447	0.497	0	1
Usually drinks alcohol with meals	251535	0.675	0.468	0	1
Drinks more than 10 years ago	274194	0.317	0.560	0	5
Drinks per week	385790	10.409	10.340	0	483
Log. drinks per week	385790	2.046	0.969	0	6
Drank yesterday	70716	0.477	0.499	0	1
Goes to the pub at least weekly	501708	0.263	0.440	0	1
Repeat Visit (Aug 2012-Jun 2013)					
Alcohol drinking frequency (6-point scale)	20333	3.115	1.461	0	5
Drinks daily	20337	0.187	0.390	0	1
Drinks at least weekly	20337	0.714	0.452	0	1
Used to alcohol drink	1363	0.479	0.500	0	1
Usually drinks alcohol with meals	12963	0.729	0.444	0	1
Drinks more than 10 years ago	10789	0.232	0.482	0	5
Drinks per week	15861	9.531	8.898	0	136
Log. drinks per week	15861	2.014	0.896	0	5
Goes to the pub at least weekly	20337	0.245	0.430	0	1
Imaging Visit (May 2014-ongoing)					
Alcohol drinking frequency (6-point scale)	23451	3.098	1.432	0	5
Drinks daily	23455	0.167	0.373	0	1
Drinks at least weekly	23455	0.716	0.451	0	1
Used to alcohol drink	1518	0.483	0.500	0	1
Usually drinks alcohol with meals	14432	0.718	0.450	0	1
Drinks more than 10 years ago	12093	0.246	0.493	0	5
Drinks per week	18278	9.532	8.668	0	106
Log. drinks per week	18278	2.028	0.881	0	5
Goes to the pub at least weekly	23455	0.264	0.441	0	1
24 hrs dietary recall (Feb 11, Jun 11, Oct 11, Apr 12)					
Drank yesterday	100604	0.505	0.500	0	1
Drank yesterday	83269	0.531	0.499	0	1
Drank yesterday	103797	0.506	0.500	0	1
Drank yesterday	100254	0.507	0.500	0	1
Mental Health Module (Jul 2016-Jul 2017)					
Alcohol drinking frequency (5-point scale)	157167	2.590	1.276	0	4
Ever had any addiction	155604	0.060	0.238	0	1
Ever addicted to alcohol	154927	0.023	0.151	0	1
Currently addicted to alcohol	149757	0.010	0.101	0	1
Typical amount of drinking	156846	1.658	1.130	0	5
Binge drinking frequency	157149	1.323	0.903	0	4

Notes. This table reports summary statistics for alcohol-related outcomes in the UK Biobank. Several measures were collected in multiple rounds of the assessment. Summary statistics are reported separately for each round.

TABLE A4.2
CORRELATION OF ALCOHOL-RELATED OUTCOMES

	Frequency (6p)	Frequency (5p)	Drinks daily	Drinks per week	Drank yesterday	Typical amount	Binge drinking	Goes to pub
Drinking Frequency (6-point)	1							
Drinking Frequency (5-point)	0.759***	1						
Drinks daily	0.658***	0.445***	1					
Drinks per week	0.576***	0.493***	0.550***	1				
Drank yesterday	0.524***	0.419***	0.416***	0.360***	1			
Typical amount of drinking	0.384***	0.440***	0.165***	0.530***	0.188***	1		
Binge drinking frequency	0.430***	0.502***	0.265***	0.565***	0.231***	0.710***	1	
Goes to the pub at least weekly	0.148***	0.145***	0.0637***	0.226***	0.0811***	0.244***	0.244***	1

Notes. This table displays the correlation of alcohol-related outcomes. Significance levels: ⁺ $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A4.3
LOCAL LICENSING POLICY IS UNCORRELATED WITH THE AVERAGE GENETIC PREDISPOSITION

	Local authorities with avg. PGS below median		Local authorities with avg. PGS above median		LHS-test Pei et al. 2019
	Mean	Standard Deviation	Mean	Standard Deviation	Coeff.
On-premise licenses (per 1,000)	1.921	0.655	1.848	1.361	0.010
No-sales premise licenses (per 1,000)	0.637	0.304	0.592	0.301	-0.001
24h supermarket licenses (per 1,000)	0.021	0.013	0.027	0.038	0.000
F-test: $\chi^2(3)$					3.500
F-test: P-value					0.321

Notes. This table reports on the balancedness of local licensing policy. Columns (1) to (4) present the summary statistics for local authorities with an average polygenic score below and above the median of all areas. Column (5) uses a left-hand-side test (Pei et al., 2019) to check whether licensing policy is predictive of the local genetic predisposition in regressions of the form $X_{i,s} = \alpha + \beta \text{PGS}_i + PC_i + f(\text{age}_i, \text{sex}_i) + \varepsilon_{i,s}$. The F-test tests whether these coefficients are jointly different from zero. Significance levels: $^+ P < 0.1$, $^* P < 0.05$, $^{**} P < 0.01$, $^{***} P < 0.001$.

TABLE A4.4
ALCOHOL CONSUMPTION AS A FUNCTION OF THE PGS AND THE NUMBER OF PUBS AND RETAILERS

	Proximity to Pubs			Proximity to Retailers				
	(1) Drinking frequency	(2) Daily drinker	(3) Drinks per week	(4) Typical amount	(5) Drinking frequency	(6) Daily drinker	(7) Drinks per week	(8) Typical amount
Pubs within 1km	0.0007+ (0.0004)	0.0008* (0.0004)	0.0006** (0.0002)	0.0005*** (0.0001)				
PGS-DPW	0.1558*** (0.0030)	0.1239*** (0.0034)	0.1603*** (0.0029)	0.1373*** (0.0044)	0.1579*** (0.0032)	0.1258*** (0.0037)	0.1621*** (0.0029)	0.1405*** (0.0045)
Pubs within 1km × PGS-DPW	-0.0003*** (0.0001)	-0.0002** (0.0001)	-0.0002+ (0.0001)	-0.0002* (0.0001)				
Retailers within 1km					0.0021 (0.0041)	0.0063* (0.0032)	0.0038 (0.0025)	0.0050*** (0.0015)
Retailers within 1km × PGS-DPW					-0.0031*** (0.0009)	-0.0025*** (0.0007)	-0.0023** (0.0007)	-0.0034*** (0.0011)
Constant	-0.0117*** (0.0029)	-0.0018 (0.0030)	-0.0209*** (0.0019)	-0.0188*** (0.0013)	-0.0090 (0.0057)	-0.0038 (0.0044)	-0.0208*** (0.0035)	-0.0215*** (0.0023)
First 40 PCs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age X Sex	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Local authority FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R2	0.105	0.044	0.136	0.116	0.105	0.044	0.136	0.116
N	486580	487147	374997	153340	486580	487147	374997	153340

Notes. This table reports the effect of the alcohol availability and genetic predisposition on four alcohol-related outcomes. Columns (1) and (5) shows the effect on the normalized drinking frequency (6-point scale), Columns (2) and (6) on the normalized probability of drinking daily, Columns (3) and (7) on the normalized log. number of alcoholic beverages consumed per week, and Columns (4) and (8) on the normalized amount of alcohol drunk in a typical drinking session. "Pubs within 1km" and "Retailers within 1km" indicates the number of pubs and retailers within 1000m of a participant's place of residence, respectively. "PGS-DPW" denotes the polygenic score for the number of drinks per week. "Pubs within 1km \times PGS-DPW" and "Retailers within 1km \times PGS-DPW" are the interactions of the alcohol availability measures with the polygenic score. All models control for genetic stratification using the first 40 principal components of the genetic data, "Age \times Sex" fixed effects, and fixed effects of the local authority. Standard errors are clustered at the local authority level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A4.5
PROBABILITY OF MOVING AS A FUNCTION
OF THE PGS AND THE NUMBER OF PUBS

	(1)	(2)	(3)	(4)
Pubs within 1km	0.042*** (0.010)	0.041*** (0.010)	0.039*** (0.010)	0.034** (0.011)
PGS-DPW	-0.210* (0.093)	-0.225* (0.090)	-0.199* (0.090)	-0.224* (0.090)
Pubs within 1km \times PGS-DPW	-0.003+ (0.002)	-0.004+ (0.002)	-0.003+ (0.002)	-0.003 (0.002)
Constant	9.286*** (0.208)	9.280*** (0.201)	9.274*** (0.203)	9.318*** (0.091)
First 40 PCs	No	Yes	Yes	Yes
Age X Sex	No	No	Yes	Yes
Local authority FE	No	No	No	Yes
R2	0.001	0.002	0.007	0.013
N	331299	331299	331296	331241

Notes. This table reports the effect of a mismatch between a participant's genetic predisposition and local alcohol availability, measured by the number of pubs in the area, on the probability of moving to another location. The dependent variable in all models takes value 100 if the participant moved since the first assessment of the UKB and 0 otherwise. The sample consists of all UKB participants that are included in at least two assessments. "Pubs within 1km" indicates the number of pubs within 1000m of a participant's place of residence. "PGS-DPW" denotes the polygenic score for the number of drinks per week. "Pubs within 1km \times PGS-DPW" is the interactions of the alcohol availability measure with the polygenic score. Column (1) includes no additional controls. Column (2) controls only for genetic stratification using the first 40 principal components of the genetic data. Column (3) controls for genetic stratification and additionally includes "Age \times Sex" fixed effects. Column (4) includes controls for genetic stratification, demographic fixed effects, and fixed effects for the local authority. Standard errors are clustered at the local authority level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A4.6
PROBABILITY OF MOVING AS A FUNCTION
OF THE PGS AND THE NUMBER OF RETAILERS

	(1)	(2)	(3)	(4)
Retailers within 1km	0.478*** (0.060)	0.458*** (0.058)	0.425*** (0.054)	0.451*** (0.049)
PGS-DPW	-0.186 ⁺ (0.100)	-0.180 ⁺ (0.095)	-0.160 ⁺ (0.095)	-0.164 ⁺ (0.095)
Retailers within 1km \times PGS-DPW	-0.056* (0.022)	-0.058** (0.022)	-0.053* (0.022)	-0.058** (0.021)
Constant	8.968*** (0.208)	8.977*** (0.200)	9.007*** (0.205)	8.972*** (0.071)
First 40 PCs	No	Yes	Yes	Yes
Age X Sex	No	No	Yes	Yes
Local authority FE	No	No	No	Yes
R2	0.001	0.002	0.007	0.013
N	331299	331299	331296	331241

Notes. This table reports the effect of a mismatch between a participant's genetic predisposition and local alcohol availability, measured by the number of retailers in the area, on the probability of moving to another location. The dependent variable in all models takes value 100 if the participant moved since the first assessment of the UKB and 0 otherwise. The sample consists of all UKB participants that are included in at least two assessments. "Retailers within 1km" indicates the number of retailers within 1000m of a participant's place of residence. "PGS-DPW" denotes the polygenic score for the number of drinks per week. "Retailers within 1km \times PGS-DPW" is the interactions of the alcohol availability measure with the polygenic score. Column (1) includes no additional controls. Column (2) controls only for genetic stratification using the first 40 principal components of the genetic data. Column (3) controls for genetic stratification and additionally includes "Age \times Sex" fixed effects. Column (4) includes controls for genetic stratification, demographic fixed effects, and fixed effects for the local authority. Standard errors are clustered at the local authority level. Significance levels: ⁺ $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A4.7
PROBABILITY OF MOVING TO ANOTHER LOCAL AUTHORITY
AS A FUNCTION OF THE PGS AND LICENSING POLICY

	(1)	(2)	(3)	(4)
On-premise licenses (per 1'000)	0.496** (0.157)	0.478** (0.151)	0.483** (0.155)	0.578 (0.671)
PGS-DPW	-0.280** (0.094)	-0.056 (0.096)	-0.043 (0.097)	-0.053 (0.086)
On-premise licenses (per 1'000) \times PGS-DPW	0.000 (0.028)	-0.009 (0.029)	-0.010 (0.030)	-0.018 (0.021)
Constant	4.597*** (0.319)	4.630*** (0.309)	4.616*** (0.318)	4.440*** (1.264)
First 40 PCs	No	Yes	Yes	Yes
Age X Sex	No	No	Yes	Yes
Local authority FE	No	No	No	Yes
R ²	0.001	0.002	0.004	0.012
N	273283	273283	273281	273236

Notes. This table reports the effect of a mismatch between a participant's genetic predisposition and local licensing policy, measured by the number of premises licensed to sell alcohol, on the probability of moving to another local authority. The dependent variable in all models takes value 100 if the participant moved to another local authority since the first assessment of the UKB and 0 otherwise. The sample consists of all UKB participants that are included in at least two assessments. "On-premise licenses (per 1,000)" indicates the number of premises per capita allowed to sell alcohol on-premise (including premises permitted to sell alcohol both on- and off-premise). "PGS-DPW" denotes the polygenic score for the number of drinks per week. "On-premise licenses (per 1,000) \times PGS-DPW" is the interactions of the policy with the polygenic score. Column (1) includes no additional controls. Column (2) controls only for genetic stratification using the first 40 principal components of the genetic data. Column (3) controls for genetic stratification and additionally includes "Age \times Sex" fixed effects. Column (4) includes controls for genetic stratification, demographic fixed effects, and fixed effects for the local authority. Standard errors are clustered at the local authority level. Significance levels: $^+P < 0.1$, $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

TABLE A4.8
LIVER DISEASES AND GENETIC PREDISPOSITION

	Alcoholic liver disease		Alcoholic fatty liver		Alcoholic hepatitis		Alcoholic liver cirrhosis		Alcoholic hepatic failure		Alc. liver disease (unspecified)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
PGS-DPW	0.054*** (0.010)	0.057*** (0.010)	0.003 (0.003)	0.004 (0.003)	0.010* (0.004)	0.010* (0.004)	0.012+ (0.007)	0.014* (0.007)	0.009* (0.004)	0.008* (0.004)	0.033*** (0.008)	0.038*** (0.009)
Drinking frequency		-0.025** (0.008)		0.001 (0.002)		-0.003 (0.004)		-0.012* (0.005)		0.001 (0.003)		-0.029*** (0.006)
Constant	0.131*** (0.008)	0.127*** (0.008)	0.006*** (0.001)	0.006*** (0.001)	0.029*** (0.003)	0.028*** (0.003)	0.052*** (0.004)	0.050*** (0.004)	0.019*** (0.003)	0.019*** (0.003)	0.068*** (0.005)	0.065*** (0.005)
First 40 PCs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age X Sex	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.001	0.001	0.000	0.000	0.001	0.001	0.001	0.001	0.000	0.000	0.001	0.001
N	380469	379567	380469	379567	380469	379567	380469	379567	380469	379567	380469	379567

Notes. This table reports the effect of a participant's genetic predisposition on the probability of having a diagnosed alcohol-related medical condition of the liver. The dependent variable in all models takes value 100 if the participant has ever been diagnosed with the condition and 0 otherwise. Columns (1) and (2) shows the effect on being diagnosed with any alcohol-related condition of the liver (ICD-10 K70), Columns (3) and (4) on alcoholic fatty liver (ICD-10 K70.0), Columns (5) and (6) on alcoholic hepatitis (ICD-10 K70.1), Columns (7) and (8) on alcoholic liver cirrhosis (ICD-10 K70.3), Columns (9) and (10) on alcoholic hepatic failure (ICD-10 K70.4), and Columns (11) and (12) on other unspecified alcohol-related liver disease (ICD-10 K70.9). "PGS-DPW" denotes the polygenic score for the number of drinks per week. "Drinking frequency" is the self-reported drinking frequency (6-point scale). All models control for genetic stratification using the first 40 principal components of the genetic data and "Age \times Sex" fixed effects. Standard errors are clustered at the local authority level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A4.9
ALCOHOL-RELATED MENTAL AND BEHAVIORAL DISORDERS AND GENETIC PREDISPOSITION

	Alcohol-induced mental and behavioral disorders		Acute intoxication with alcohol		Harmful use of alcohol		Alcoholic dependence syndrome		Alcoholic withdrawal state		Alcoholic psychotic disorder		Alc. mental disorder (unspecified)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
PGS-DPW	0.138*** (0.021)	0.141*** (0.022)	0.038*** (0.011)	0.036** (0.011)	0.019** (0.007)	0.016* (0.007)	0.064*** (0.014)	0.066*** (0.014)	0.070*** (0.013)	0.073*** (0.013)	0.002 (0.002)	0.003 (0.002)	0.004 (0.003)	0.004 (0.003)
Drinking frequency		-0.084*** (0.014)		-0.017* (0.008)		-0.004 (0.005)		-0.044*** (0.009)		-0.048*** (0.010)		-0.005* (0.002)		-0.001 (0.002)
Constant	0.338*** (0.017)	0.323*** (0.016)	0.116*** (0.008)	0.110*** (0.007)	0.046*** (0.004)	0.044*** (0.004)	0.117*** (0.010)	0.110*** (0.010)	0.145*** (0.009)	0.136*** (0.009)	0.007*** (0.001)	0.006*** (0.001)	0.007*** (0.002)	0.006*** (0.002)
First 40 PCs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age X Sex	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.003	0.003	0.001	0.001	0.001	0.001	0.002	0.002	0.002	0.002	0.001	0.001	0.001	0.001
N	380469	379567	380469	379567	380469	379567	380469	379567	380469	379567	380469	379567	380469	379567

Notes. This table reports the effect of a participant's genetic predisposition on the probability of having a diagnosed alcohol-related mental or behavioral disorder. The dependent variable in all models takes value 100 if the participant has ever been diagnosed with the condition and 0 otherwise. Columns (1) and (2) shows the effect on being diagnosed with any alcohol-related mental or behavioral disorder (ICD-10 F10), Columns (3) and (4) on acute intoxication with alcohol (ICD-10 F10.0), Columns (5) and (6) on harmful use of alcohol (ICD-10 F10.1), Columns (7) and (8) on alcoholic dependence syndrome (ICD-10 F10.2), Columns (9) and (10) on alcoholic withdrawal state with possible delirium (ICD-10 F10.3 and F10.4), Columns (11) and (12) on alcoholic psychotic disorder (ICD-10 F10.5), and Columns (13) and (14) on other unspecified alcohol-related mental and behavioral disorders (ICD-10 K70.9). "PGS-DPW" denotes the polygenic score for the number of drinks per week. "Drinking frequency" is the self-reported drinking frequency (6-point scale). All models control for genetic stratification using the first 40 principal components of the genetic data and "Age × Sex" fixed effects. Standard errors are clustered at the local authority level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

TABLE A4.10
OTHER ALCOHOL-RELATED DISEASES AND GENETIC PREDISPOSITION

	Degeneration of nervous system due to alcohol		Alcoholic gastritis		Alcoholic pancreatitis (acute)		Alcoholic pancreatitis (chronic)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PGS-DPW	0.005 (0.003)	0.004 (0.003)	0.014** (0.004)	0.013** (0.004)	0.005 (0.003)	0.004 (0.003)	0.022*** (0.005)	0.020*** (0.005)
Drinking frequency		-0.002 (0.002)		-0.005 (0.004)		0.005** (0.002)		-0.008+ (0.004)
Constant	0.008*** (0.002)	0.007*** (0.001)	0.024*** (0.002)	0.022*** (0.002)	0.009*** (0.001)	0.010*** (0.001)	0.031*** (0.003)	0.029*** (0.003)
First 40 PCs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Age X Sex	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R ²	0.000	0.000	0.001	0.001	0.000	0.000	0.001	0.001
N	380469	379567	380469	379567	380469	379567	380469	379567

Notes. This table reports the effect of a participant's genetic predisposition on the probability of having another diagnosed alcohol-related condition. The dependent variable in all models takes value 100 if the participant has ever been diagnosed with the condition and 0 otherwise. Columns (1) and (2) shows the effect on being diagnosed with degeneration of the nervous system due to alcohol (ICD-10 G31.2), Columns (3) and (4) on alcoholic gastritis (ICD-10 K29.2), Columns (5) and (6) on acute alcoholic pancreatitis (ICD-10 K85.2), and Columns (7) and (8) on chronic alcoholic pancreatitis (ICD-10 K86.0). "PGS-DPW" denotes the polygenic score for the number of drinks per week. "Drinking frequency" is the self-reported drinking frequency (6-point scale). All models control for genetic stratification using the first 40 principal components of the genetic data and "Age \times Sex" fixed effects. Standard errors are clustered at the local authority level. Significance levels: + $P < 0.1$, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

REFERENCES

- PEI, Z., J.-S. PISCHKE, AND H. SCHWANDT (2019): “Poorly measured confounders are more useful on the left than on the right,” *Journal of Business & Economic Statistics*, 37, 205–216.
- UKB (2006): “UK Biobank: Protocol for a large-scale prospective epidemiological resource,” *Protocol No: UKBB-PROT-09-06 (Main Phase)*.

Curriculum Vitae

Personal details

Name:	Christian Lukas Zünd
Date of Birth:	27 June 1990

Education

August 2014 – April 2020:	Doctoral Studies in Economics, University of Zurich, Department of Economics, Zurich Graduate School of Economics
October 2018 – August 2019:	Visiting PhD Student, University of Oxford, Centre for Experimental Social Science and Department of Economics
October 2013 – August 2014:	Master of Philosophy in Economic Research, Univer- sity of Cambridge (Trinity College), Department of Economics
September 2010 – September 2013:	Bachelor of Arts in Economics, University of Zurich, Department of Economics